

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:38:32 ; Search time 59 Seconds
(without alignments)
2078.400 Million cell updates/sec

Title: US-09-985-689A-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSYGLY.....EVOQNVNVPQTFSLAIVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	2247	100.0	434	5	AAW50080	AAW50080	Bacillus
2	2247	100.0	640	2	AAAY17090	AAAY17090	Bacillus
3	2242	99.8	640	2	AAAY17091	AAAY17091	Bacillus
4	2191	97.5	434	5	AAW50081	AAW50081	Bacillus
5	2183	97.2	639	2	AAAY17089	AAAY17089	Bacillus
6	2155	95.9	639	2	AAAY17087	AAAY17087	An alkali
7	2155	95.9	640	2	AAAY17088	AAAY17088	An alkali
8	2143	95.4	434	5	AAW50085	AAW50085	Bacillus
9	2125.5	94.6	433	5	AAW50086	AAW50086	Bacillus
10	2125.5	94.6	641	2	AAW89547	AAW89547	Bacillus
11	2071	93.2	434	5	AAW50090	AAW50090	Bacillus
12	1998.5	88.9	433	5	AAW50084	AAW50084	Bacillus
13	1994.5	88.8	433	5	AAW50082	AAW50082	Bacillus
14	1989.5	88.5	636	2	AAW89548	AAW89548	Bacillus
15	1987.5	88.5	433	5	AAW50083	AAW50083	Bacillus
16	1986.5	88.4	433	2	AAW61495	AAW61495	Modified
17	1986.5	88.4	433	2	AAW61495	AAW61495	Modified
18	1986.5	88.4	433	2	AAW95698	AAW95698	Bacillus
19	1986.5	88.4	433	3	AAW63207	AAW63207	Amino aci
20	1986.5	88.4	433	3	AAW44619	AAW44619	Bacillus
21	1581.5	70.4	345	2	AAW62230	AAW62230	Subtilase
22	1581.5	70.4	345	2	AAW21654	AAW21654	Subtilase
23	452.5	20.1	659	2	AAW24121	AAW24121	Thermococ
24	452.5	20.1	659	2	AAW94840	AAW94840	WO9856926
25	414	18.4	412	2	AAW94836	AAW94836	Hyperther

26	414	18.4	522	2	AAW24122	AAW24122	Pyrococcus
27	414	18.4	522	2	AAW94838	AAW94838	Hyperther
28	414	18.4	654	2	AAW24129	AAW24129	Pyrococcus
29	414	18.4	654	2	AAW94841	AAW94841	Hyperther
30	401	17.8	659	2	AAW24123	AAW24123	Protease.
31	372	16.6	545	4	ABB09483	ABB09483	T. Yonsei
32	366	16.3	1079	4	ABB81180	ABB81180	Transglut
33	366	16.3	1079	6	ABU07391	ABU07391	Foreign p
34	346	15.4	520	2	AAW13666	AAW13666	Streptomy
35	346	15.4	734	2	AAW13667	AAW13667	Streptomy
36	346	15.4	823	2	AAW13668	AAW13668	DhpA-mel
37	338	15.0	1237	6	ABU11343	ABU11343	Protein e
38	310.5	13.8	806	2	AAW27481	AAW27481	RP-III re
39	306.5	13.6	519	6	ABP76735	ABP76735	Streptomy
40	306.5	13.6	19938	6	ABP76678	ABP76678	Streptomy
41	304.5	13.6	903	2	AAW87007	AAW87007	Hyperther
42	304.5	13.6	1398	2	AAW87008	AAW87008	Protease.
43	304.5	13.6	1398	2	AAW24124	AAW24124	Pyrococcus
44	304.5	13.6	1398	2	AAW94839	AAW94839	WO9856926
45	301.5	13.4	699	2	AAW08471	AAW08471	P. balust

ALIGNMENTS

RESULT 1
AAW50080
ID AAW50080 standard; protein; 434 AA.
XX
AC AAW50080;
XX
DT 12-AUG-2002 (first entry)
XX
DE Bacillus sp KSM-KP43 alkaline protease protein fragment.
XX
KW Alkaline protease; detergent; laundry; bleaching; dishwasher.
XX
OS Bacillus sp.
XX
PN EP1209233-A2.
XX
PD 29-MAY-2002.
XX
PF 22-NOV-2001; 2001EP-00127851.
XX
PR 22-NOV-2000; 2000JP-00355156.
PR 12-APR-2001; 2001JP-00114048.
XX
PA (KAOS) KAO CORP.
PI Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
PI Okuda M, Saeki K;
XX
DR WPI; 2002-437518/47.
XX
FT New modified alkaline proteases useful in detergent compositions.
XX
PS Claim 1; Page 10-11; 25pp; English.
XX
CC This invention describes novel Bacillus sp. alkaline proteases useful in
CC detergent compositions, especially in laundry, bleaching or automatic
CC dishwasher detergents, the novel proteases have an increased detergency *
CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
CC sequence represents a fragment of the alkaline protease KP43 from
CC Bacillus sp strain KSM-KP43 which is used to create the modified protease
CC represented in AAW50090
XX
SQ Sequence 434 AA;

Query Match 100.0%; Score 2247; DB 5; Length 434;
Best Local Similarity 100.0%; Pred. No. 3.9e-159;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
 Db 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
 QY 61 NANDTNGHGHVAGSVLNGSTNKGAPQANLVFQSIIMDSGGGLGGLPSNLQTLFQAYS 120
 Db 61 NANDTNGHGHVAGSVLNGSTNKGAPQANLVFQSIIMDSGGGLGGLPSNLQTLFQAYS 120
 QY 121 AGARIHTNSWGAANVAGYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 Db 121 AGARIHTNSWGAANVAGYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 QY 181 TVGATENLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVAPGTFILSARSSLPDSSF 240
 Db 181 TVGATENLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVAPGTFILSARSSLPDSSF 240
 QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 Db 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 QY 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 Db 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 QY 361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGGRNNVNFVINAPOSQGTYYIEVOAYN 420
 Db 361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGGRNNVNFVINAPOSQGTYYIEVOAYN 420
 QY 421 VPVGPQTFSLAIVN 434
 Db 421 VPVGPQTFSLAIVN 434

RESULT 2

AAV17090
 ID AAV17090 standard; protein; 640 AA.

XX AAV17090;
 AC
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE Bacillus alkaline protease.
 XX
 KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent.
 XX
 OS Bacillus sp.
 XX
 PN WO9918218-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX
 PR 07-OCT-1997; 97JP-00274570.
 XX
 PA (KAOS) KAO CORP.
 XX
 PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX
 DR WPI; 1999-287736/27.
 DR N-PSDB; AAX37276.
 XX
 XX Alkali protease from Bacillus used in washing powders.
 PT
 XX
 XX Disclosure; Page 58-63; 71pp; Japanese.

XX
 CC The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it

CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease. (Updated
 CC on 20-MAR-2003 to correct DR field.)
 XX

SQ Sequence 640 AA;

Query Match 100.0%; Score 2247; DB 2; Length 640;
 Best Local Similarity 100.0%; Pred. No. 6.5e-159;
 Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
 Db 207 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266
 QY 61 NANDTNGHGHVAGSVLNGSTNKGAPQANLVFQSIIMDSGGGLGGLPSNLQTLFQAYS 120
 Db 267 NANDTNGHGHVAGSVLNGSTNKGAPQANLVFQSIIMDSGGGLGGLPSNLQTLFQAYS 326
 QY 121 AGARIHTNSWGAANVAGYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 Db 327 AGARIHTNSWGAANVAGYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
 QY 181 TVGATENLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVAPGTFILSARSSLPDSSF 240
 Db 387 TVGATENLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVAPGTFILSARSSLPDSSF 446
 QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 Db 447 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 506
 QY 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 Db 507 PNGNQGWGRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 566
 QY 361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGGRNNVNFVINAPOSQGTYYIEVOAYN 420
 Db 567 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGGRNNVNFVINAPOSQGTYYIEVOAYN 626
 QY 421 VPVGPQTFSLAIVN 434
 Db 627 VPVGPQTFSLAIVN 640

RESULT 3

AAV17091
 ID AAV17091 standard; protein; 640 AA.

XX AAV17091;
 AC
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE Bacillus alkaline protease.
 XX
 KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent.
 XX
 OS Bacillus sp.
 XX
 PN WO9918218-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX
 PR 07-OCT-1997; 97JP-00274570.

XX PA (KAOS) KAO CORP.
XX PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
XX PI Shikata S, Nomura M;
XX DR WPI; 1999-287736/27.
XX DR N-PSDB; AAX37279.
XX PT Alkali protease from Bacillus used in washing powders.
XX PS Disclosure; Page 63-68; 7lpp; Japanese.
XX CC The invention relates to alkaline proteases produced by strains of
XX CC Bacillus. The proteases ability to digest casein is not inhibited by
XX CC oleic acid and they have a high stability to oxidising agents. The
XX CC alkaline protease of the invention has the following properties: (a) it
XX CC is active over the pH range 4-13 and has at least 80% of its optimum
XX CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
XX CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
XX CC its ability to digest casein is not inhibited by oleic acid; (e) it has
XX CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
XX CC used as enzymes in washing compositions for use in automatic dishwashers
XX CC and for washing clothes. The stability to oxidising agents allows the
XX CC enzyme to be an effective component of washing compositions including
XX CC bleaches. The present sequence represents an alkaline protease. (Updated
XX CC on 20-MAR-2003 to correct DR field.)
XX SQ Sequence 640 AA;
Query Match 99.8%; Score 2242; DB 2; Length 640;
Best Local Similarity 99.8%; Pred. No. 1.5e-158;
Matches 433; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 NDVARGIVKADVAQSSYGLYGQGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
DB 207 NDVARGIVKADVAQSSYGLYGQGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAQANLVFQSIIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 267 NANDTNGHGHVAGSVLNGSTNKGMAQANLVFQSIIMDSGGGLGGLPSNLQTLFSQAYS 326
QY 121 AGARIHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 327 AGARIHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
QY 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVWAPGTFILSARSSLPDSSF 240
DB 387 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVWAPGTFILSARSSLPDSSF 446
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
DB 447 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 506
QY 301 PNGNQGWRVTLDKSLNVAAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
DB 507 PNGNQGWRVTLDKSLNVAAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 566
QY 361 SVTLVNDLVLITAPNGTQYVGNDFTPSYNDWGRNNVNFVFNAPQSGTITIEVQAYN 420
DB 567 SVTLVNDLVLITAPNGTQYVGNDFTPSYNDWGRNNVNFVFNAPQSGTITIEVQAYN 626
QY 421 VPVGQPTFSLAIVN 434
DB 627 VPVGQPTFSLAIVN 640
RESULT 4
AAM50081
ID AAM50081 standard; protein; 434 AA.
XX AC
XX AC AAM50081;
XX XX

42x

DT 12-AUG-2002 (first entry)
XX DE Bacillus sp-KSM-KP9860 alkaline protease protein fragment.
XX KW Alkaline protease; detergent; laundry; bleaching; dishwasher.
XX OS Bacillus sp.
XX PN EP1209233-A2.
XX PD 29-MAY-2002.
XX XX 22-NOV-2001; 2001EP-00127851.
XX PR 22-NOV-2000; 2000JP-00355166.
XX PR 12-APR-2001; 2001JP-00114048.
XX PA (KAOS) KAO CORP.
XX XX Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
XX PI Okuda M, Saeki K;
XX XX WPI; 2002-437518/47.
XX DR New modified alkaline proteases useful in detergent compositions.
XX PT Claim 5; Page 12-13; 25pp; English.
XX PS This invention describes novel Bacillus sp. alkaline proteases useful in
XX CC detergent compositions, especially in laundry, bleaching or automatic
XX CC dishwasher detergents. The novel proteases have an increased detergency %
XX CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
XX CC sequence represents a fragment of the alkaline protease KP9860 from
XX CC Bacillus sp strain KSM-KP9860 described in the method of the invention
XX SQ Sequence 434 AA;
Query Match 97.5%; Score 2191; DB 5; Length 434;
Best Local Similarity 96.5%; Pred. No. 5.8e-155;
Matches 419; Conservative 13; Mismatches 2; Indels 0; Gaps 0;
QY 1 NDVARGIVKADVAQSSYGLYGQGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
DB 1 NDVARGIVKADVAQSSYGLYGQGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAQANLVFQSIIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NANDTNGHGHVAGSVLNGSTNKGMAQANLVFQSIIMDSGGGLGGLPSNLQTLFSQAYS 120
QY 121 AGARIHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 121 AGARIHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVWAPGTFILSARSSLPDSSF 240
DB 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVWAPGTFILSARSSLPDSSF 240
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
DB 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
QY 301 PNGNQGWRVTLDKSLNVAAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
DB 301 PNGNQGWRVTLDKSLNVAAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
QY 361 SVTLVNDLVLITAPNGTQYVGNDFTPSYNDWGRNNVNFVFNAPQSGTITIEVQAYN 420
DB 361 SVTLVNDLVLITAPNGTQYVGNDFTPSYNDWGRNNVNFVFNAPQSGTITIEVQAYN 420
QY 421 VPVGQPTFSLAIVN 434
DB 421 VPVGQPTFSLAIVN 434

42x

RESULT 5
 AAY17089
 ID AAY17089 standard; protein; 639 AA.
 XX
 AC AAY17089;
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 XX Bacillus alkaline protease.
 DE Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 XX washing composition; oxidising agent.
 KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent.
 OS Bacillus sp.
 XX
 XX WO9918218-Al.
 XX
 XX 15-APR-1999.
 XX
 XX 07-OCT-1998; 98WO-JP004528.
 XX
 XX 07-OCT-1997; 97JP-00274570.
 XX
 XX (KAOS) KAO CORP.
 XX
 XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX
 XX WPI; 1999-287736/27.
 DR N-PSDB; AAX37277.
 XX
 XX Alkali protease from Bacillus used in washing powders.
 PT
 XX Disclosure; Page 53-59; 71pp; Japanese.
 PS
 XX The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease. (Updated
 CC on 20-MAR-2003 to correct DR field.)
 XX
 XX SQ
 Query Match 97.2%; Score 2183; DB 2; Length 639;
 Best Local Similarity 96.3%; Pred. No. 3.9e-154;
 Matches 418; Conservative 13; Mismatches 3; Indels 0; Gaps 0;
 1 NDVARGIVKADVAQSSYGLYGGQQIVAVADTGLDGRNDSMHEAPRGKITLYALGRTN 60
 206 NDVARGIVKADVAQSSYGLYGGQQIVAVADTGLDGRNDSMHEAPRGKITLYALGRTN 265
 61 NNDTNGHGHVAGSVLNGSTNKGMAQANLVFQISMDSGGLGGLPSNLQTLFSQAYS 120
 266 NNDTNGHGHVAGSVLNGATNKGMAQANLVFQISMDSGGLGGLPSNLQTLFSQAFS 325
 121 AGARIHTNSGAAVNGAYTTDSRDYVRKNDMTILFAAGNEGPGNGTISAPGTAKNAI 180
 326 AGARIHTNSGAAVNGAYTTDSRDYVRKNDMTILFAAGNERPNGGTISAPGTAKNAI 385
 181 TVGATENLRFSFGYADNINHAQFSRSGPTKDGRIKPDVMAPTGTFILSARSSLAPDSF 240
 386 TVGATENLRFSFGYADNINHAQFSRSGPTKDGRIKPDVMAPTGTFILSARSSLAPDSF 445

241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSILLKALINGRADIGLY 300
 446 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSILLKALINGRADIGLY 505
 301 PNGNQGWGRVTLDKSLNVAAYVNESSSLSTSOKATYSFTATAGKPLKISLWSDAPASTTA 360
 506 PNGNQGWGRVTLDKSLNVAAYVNESSALSTSOKATYFTATAGKPLKISLWSDAPASTTA 565
 361 SVTLVNDLDLIVITAPNGTQVVGNDFTSPYNDNDGRNVNENFINAPQSGTYTIEVQAYN 420
 566 SVTLVNDLDLIVITAPNGTRYVGNDFSPFNDNDRNVNENFINPQSGTYTIEVQAYN 625
 421 VPVGPOTFSLAIYN 434
 526 VPVGPOTFSLAIYN 639

RESULT 6
 AAY17087
 ID AAY17087 standard; protein; 639 AA.
 XX
 AC AAY17087;
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 XX An alkaline protease sequence from Bacillus species.
 DE Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent.
 KW
 OS Bacillus sp.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 1..639
 FT /note= "all residues indicated as Xaa are arbitrary amino
 FT acids"
 XX
 XX WO9918218-Al.
 XX
 XX 15-APR-1999.
 XX
 XX 07-OCT-1998; 98WO-JP004528.
 XX
 XX 07-OCT-1997; 97JP-00274570.
 XX
 XX (KAOS) KAO CORP.
 XX
 XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX
 XX WPI; 1999-287736/27.
 DR N-PSDB; AAX37277.
 XX
 XX Alkali protease from Bacillus used in washing powders.
 PT
 XX Disclosure; Page 53-59; 71pp; Japanese.
 PS
 XX The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease. (Updated
 CC on 20-MAR-2003 to correct DR field.)
 XX
 XX SQ


```

XX SQ Sequence 639 AA;
Query Match 95.9%; Score 2155; DB 2; Length 639;
Best Local Similarity 96.3%; Pred. No. 4.7e-152;
Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 1 NDVARGIVKADVQSSYGLYGGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 206 NDVARGIVKADVQSSYGLYGGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 265
QY 61 NANDTNGHGTTHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 266 NANDTNGHGTTHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 325
QY 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTTISAPGTAKNAI 180
Db 326 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTTISAPGTAKNAI 385
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIKPDVWAPGTTFILSARSSLAPDSFF 240
Db 386 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIKPDVWAPGTTFILSARSSLAPDSFF 445
QY 241 WANHDSKYAYMGTSMATPIVAGNVAOLREHFVNKRGITPKPSLLKALIAAGADIGLY 300
Db 446 WANHDSKYAYMGTSMATPIVAGNVAOLREHFVNKRGITPKPSLLKALIAAGADIGLY 505
QY 301 PNGNQGWGRVTLDKSLNVAVYVNESSLSSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 360
Db 506 PNGNQGWGRVTLDKSLNVAVYVNESSLSSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 565
QY 361 SVTLVNDLDLIVITAPNGTQYVGNDFTPSYNDNDGRNNVENFINAPQSGTYTIEVQAYN 420
Db 566 SVTLVNDLDLIVITAPNGTQYVGNDFTPSYNDNDGRNNVENFINAPQSGTYTIEVQAYN 625
QY 421 VPVGPQTFSLAIVN 434
Db 626 VPVGPQTFSLAIVN 639

RESULT 7
AAAY17088
ID AAAY17088 standard; protein; 640 AA.
XX AC AAAY17088;
XX DT 20-MAR-2003 (revised)
XX DT 21-JUL-1999 (first entry)
XX DE An alkaline protease sequence from Bacillus species.
XX KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
XX KW washing composition; oxidising agent.
XX OS Bacillus sp.
XX FH Key Location/Qualifiers
XX FT Misc-difference 1..640
XX FT /note= "all residues indicated as Xaa are arbitrary amino
XX FT acids"
XX PN W09918218-A1.
XX PD 15-APR-1999.
XX PF 07-OCT-1998; 98WO-JP004528.
XX PR 07-OCT-1997; 97JP-00274570.
XX PA (KAOS ) KAO CORP.
XX PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
XX PI Shikata S, Nomura M;

```

```

XX DR WPI; 1999-287736/27.
XX N-PSDB; AAX37278.
XX PT Alkali protease from Bacillus used in washing powders.
XX PS Claim 3; Page 50-53; 71pp; Japanese.
XX CC The invention relates to alkaline proteases produced by strains of
XX CC Bacillus. The proteases ability to digest casein is not inhibited by
XX CC oleic acid and they have a high stability to oxidising agents. The
XX CC alkaline protease of the invention has the following properties: (a) it
XX CC is active over the pH range 4-13 and has at least 80% of its optimum
XX CC stable over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
XX CC its ability to digest casein is not inhibited by oleic acid; (e) it has
XX CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
XX CC used as enzymes in washing compositions for use in automatic dishwashers
XX CC and for washing clothes. The stability to oxidising agents allows the
XX CC enzyme to be an effective component of washing compositions including
XX CC bleaches. The present sequence represents an alkaline protease of the
XX CC invention. (Updated on 20-MAR-2003 to correct DR field.)
XX SQ Sequence 640 AA;
Query Match 95.9%; Score 2155; DB 2; Length 640;
Best Local Similarity 96.3%; Pred. No. 4.7e-152;
Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 1 NDVARGIVKADVQSSYGLYGGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 207 NDVARGIVKADVQSSYGLYGGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266
QY 61 NANDTNGHGTTHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 267 NANDTNGHGTTHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 326
QY 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTTISAPGTAKNAI 180
Db 327 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTTISAPGTAKNAI 386
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIKPDVWAPGTTFILSARSSLAPDSFF 240
Db 387 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIKPDVWAPGTTFILSARSSLAPDSFF 446
QY 241 WANHDSKYAYMGTSMATPIVAGNVAOLREHFVNKRGITPKPSLLKALIAAGADIGLY 300
Db 447 WANHDSKYAYMGTSMATPIVAGNVAOLREHFVNKRGITPKPSLLKALIAAGADIGLY 506
QY 301 PNGNQGWGRVTLDKSLNVAVYVNESSLSSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 360
Db 507 PNGNQGWGRVTLDKSLNVAVYVNESSLSSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 566
QY 361 SVTLVNDLDLIVITAPNGTQYVGNDFTPSYNDNDGRNNVENFINAPQSGTYTIEVQAYN 420
Db 567 SVTLVNDLDLIVITAPNGTQYVGNDFTPSYNDNDGRNNVENFINAPQSGTYTIEVQAYN 626
QY 421 VPVGPQTFSLAIVN 434
Db 627 VPVGPQTFSLAIVN 640

RESULT 8
AAAM50085
ID AAAM50085 standard; protein; 434 AA.
XX AC AAAM50085;
XX DT 12-AUG-2002 (first entry)
XX DE Bacillus sp alkaline protease protein A-1 fragment.
XX KW Alkaline protease; detergent; laundry; bleaching; dishwasher.

```

```

XX OS Bacillus sp.
XX PN EP1209233-A2.
XX PD 29-MAY-2002.
XX XX
XX XX 22-NOV-2001; 2001EP-00127851.
XX PF 22-NOV-2000; 2000JP-00355166.
XX PR 12-APR-2001; 2001JP-00114048.
XX XX
XX PA (KAOS ) KAO CORP.
XX XX
XX PI Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
XX PI Okuda M, Saeki K;
XX XX
XX DR WPI; 2002-437518/47.
XX XX
XX PT New modified alkaline proteases useful in detergent compositions.
XX PS Claim 5; Page 18-19; 25pp; English.
XX XX
XX CC This invention describes novel Bacillus sp. alkaline proteases useful in
XX CC detergent compositions, especially in laundry, bleaching or automatic
XX CC dishwasher detergents. The novel proteases have an increased detergency &
XX CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
XX CC sequence represents a fragment of the alkaline protease A-1 from Bacillus
XX CC sp NCIB12289 described in the method of the invention
XX XX
XX SQ Sequence 434 AA;
XX
Query Match 95.4%; Score 2143; DB 5; Length 434;
Best Local Similarity 93.5%; Pred. No. 2.2e-151;
Matches 406; Conservative 19; Mismatches 9; Indels 0; Gaps 0;
XX
QY 1 NDVARGIVKADVAQSSYGLYGGQIIVAVADTGLDTCGRNDSMHEAFRGKITAIYALGRIN 60
DB 1 NDVARGIVKADVAQSSYGLYGGQIIVAVADTGLDTCGRNDSMHEAFRGKITAIYALGRIN 60
QY 61 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
QY 121 AGARIHTNSWGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 121 AGARIHTNSWGAPVNGAYTTDSRNVDYVRKNDMAVLFAAGNEGPNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAFGTIFLSARSSLAPDSF 240
DB 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAFGTIFLSARSSLAPDSF 240
QY 241 WANHDSKYATMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLY 300
DB 241 WANHDSKYATMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAATDILGY 300
QY 301 PNGNQGWGRVTLDKSLNVAFYVNESSLSSTSQATYSFTATAGKPLKISLVWSDAPASTTA 360
DB 301 PNGNQGWGRVTLDKSLNVAFYVNESSLSSTSQATYSFTATAGKPLKISLVWSDAPASTTA 360
QY 361 SVTLVNDLDELIVITAPNGTQVGNDFTSYNDNDWGRNENVENFINAPQSGTITIEVQAYN 420
DB 361 SVTLVNDLDELIVITAPNGTQVGNDFTSYNDNDWGRNENVENFINAPQSGTITIEVQAYN 420
QY 421 VPVGPQTFFSLAIYN 434
DB 421 VPVGPQTFFSLAIYN 434

```

RESULT 9
AAM50086
ID AAM50086 standard; protein; 433 AA.
XX

```

AAM50086;
12-AUG-2002 (first entry)
Bacillus sp alkaline protease protein A-2 fragment.
Alkaline protease; detergent; laundry; bleaching; dishwasher.
Bacillus sp.
EP1209233-A2.
29-MAY-2002.
22-NOV-2001; 2001EP-00127851.
22-NOV-2000; 2000JP-00355166.
12-APR-2001; 2001JP-00114048.
(KAOS ) KAO CORP.
Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
Okuda M, Saeki K;
WPI; 2002-437518/47.
New modified alkaline proteases useful in detergent compositions.
Claim 5; Page 20-21; 25pp; English.
This invention describes novel Bacillus sp. alkaline proteases useful in
detergent compositions, especially in laundry, bleaching or automatic
dishwasher detergents. The novel proteases have an increased detergency &
(34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
sequence represents a fragment of the alkaline protease A-2 from Bacillus
sp NCIB12513 described in the method of the invention
XX
XX SQ Sequence 433 AA;
XX
Query Match 94.6%; Score 2125.5; DB 5; Length 433;
Best Local Similarity 93.5%; Pred. No. 4.4e-150;
Matches 406; Conservative 19; Mismatches 8; Indels 1; Gaps 1;
QY 1 NDVARGIVKADVAQSSYGLYGGQIIVAVADTGLDTCGRNDSMHEAFRGKITAIYALGRIN 60
DB 1 NDVARGIVKADVAQSNFGLYGGQIIVAVADTGLDTCGRNDSMHEAFRGKITAIYALGRIN 60
QY 61 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NANTNGHGHVAGSVLNG-ATNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 119
QY 121 AGARIHTNSWGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 120 AGARIHTNSWGAPVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAFGTIFLSARSSLAPDSF 240
DB 180 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAFGTIFLSARSSLAPDSF 239
QY 241 WANHDSKYATMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLY 300
DB 240 WANHDSKYATMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADVGLGF 299
QY 301 PNGNQGWGRVTLDKSLNVAFYVNESSLSSTSQATYSFTATAGKPLKISLVWSDAPASTTA 360
DB 300 PNGNQGWGRVTLDKSLNVAFYVNETSPLSTSQATYSFTAQAGKPLKISLVSDAPGSTTA 359
QY 361 SVTLVNDLDELIVITAPNGTQVGNDFTSYNDNDWGRNENVENFINAPQSGTITIEVQAYN 420
DB 360 SVTLVNDLDELIVITAPNGTKYVGNDFTFAPYDNNWGRNENVENFINAPQSGTITIEVQAYN 419
QY 421 VPVGPQTFFSLAIYN 434
XX

```

```
Db 420 VPVSPQTFSLAIVH 433
RESULT 10
AAW89547
ID AAW89547 standard; protein; 641 AA.
XX
XX AAW89547;
XX
XX 12-APR-1999 (first entry)
XX
XX Bacillus JPI170 protease.
XX
XX Protease; detergent; surfactant; leather processing; debittering;
KW flavour.
XX
XX Bacillus sp.
XX
XX Key Location/Qualifiers
XX Peptide 1.33
XX /note= "signal peptide"
XX Region 34..208
XX /note= "prepro region"
XX Protein 209..641
XX /note= "mature protein"
XX
XX W09856927-A2.
XX
XX 17-DEC-1998.
XX
XX 09-JUN-1998; 98WO-US012005.
XX
XX 12-JUN-1997; 97US-00873479.
XX
XX (NOVO ) NOVO NORDISK BIOTECH INC.
XX
XX Sloma A, Christianson L;
XX
XX WPI; 1999-080908/07.
XX N-PSDB; AAW82382.
XX
XX Novel protease from Bacillus subtilis LC20 - useful in laundry and
XX dishwashing detergents and for leather processing.
XX
XX Claim 7; Page 53-54; 77pp; English.
XX
XX This is the amino acid sequence of a novel protease of Bacillus sp. JPI170
XX (NCIB 12513), as deduced from the nucleotide sequence of an isolated gene
XX (see AAW82382). The entire protein, including the signal peptide and
XX prepro region, has 77% identity to alkaline protease Y (see AAW89548)
XX from Bacillus. The invention provides vectors, recombinant host cells and
XX methods for the recombinant production of the protease. The protease is
XX used in laundry and dishwashing detergents, for institutional and
XX industrial cleaning, and for leather processing, as well as for
XX debittering and enhancing the degree of hydrolysis of protein
XX hydrolysates, for flavour development through hydrolysis of proteins.
XX degradation of undesired peptides and in enzymatic synthesis of peptides.
XX It has enhanced stability towards oxidation under alkaline conditions,
XX e.g. towards bleaching agents of the peroxy type. The invention also
XX provides mutant cells in which the protease activity is diminished. Such
XX cells can be used for the production of heterologous recombinant proteins
XX
XX Sequence 641 AA;
XX
Query Match 94.6%; Score 2125.5; DB 2; Length 641;
Best Local Similarity 93.5%; Pred. No. 7.5e-150;
Matches 406; Conservative 19; Mismatches 8; Indels 1; Gaps 1;
XX
XX 1 NDVARGIVKADVAQSYGLYGGQIVAVADTGLDTRGNDSSMHEAFRGKITALYALGRTN 60
XX
XX 209 NDVARGIVKADVAQNFGLYGGQIVAVADTGLDTRGNDSSMHEAFRGKITALYALGRTN 268
XX
XX 61 NANTDNGHGHVAGSVLGNSTNKGMAPQANLVFQSIQNDGGGLGLPSNLQTLFSQAYS 120
```

FT Misc-difference 124 /note= "as claimed in Claim 3"

FT /label= a,k

FT /note= "as claimed in Claim 3"

FT Misc-difference 136

FT /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a

FT /note= "as claimed in Claim 3"

FT Misc-difference 138

FT /label= y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c

FT /note= "as claimed in Claim 3"

FT Misc-difference 148

FT /label= Y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c

FT /note= "as claimed in Claim 3"

FT Misc-difference 193

FT /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a

FT /note= "as claimed in Claim 3"

FT Misc-difference 195

FT /label= Y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c

FT /note= "as claimed in Claim 3"

FT Misc-difference 205

FT /label= Y,w,a,d,e,t,v,l,i,h,s,k,q,m,c

FT /note= "as claimed in Claim 3"

FT Misc-difference 247

FT /label= w,f,a,r,e,t,v,l,i,h,s,e,m,c

FT /note= "as claimed in Claim 3"

FT Misc-difference 256

FT /label= OTHER= a,s,e,v,l,r,e,d

FT /note= "OTHER= deleted residue. Specifically described in Claim 1"

FT Misc-difference 257

FT /label= v,i

FT /note= "as claimed in Claim 3"

FT Misc-difference 264

FT /label= e,d,s,q,a,t,l,m,c,v,g,i

FT /note= "as claimed in Claim 3"

FT Misc-difference 342

FT /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a

FT /note= "as claimed in Claim 3"

FT Misc-difference 369

FT /label= OTHER, d

FT /note= "OTHER= deleted residue. Specifically described in Claim 1"

FT

PN EP1209233-A2.

XX 29-MAY-2002.

XX 22-NOV-2001; 2001BP-00127851.

XX 22-NOV-2000; 2000JP-00355166.

PR 12-APR-2001; 2001JP-00114048.

XX

PA (KAOS) KAO CORP.

XX Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;

PI Okuda M, Saeki K;

XX WPI; 2002-437518/47.

XX New modified alkaline proteases useful in detergent compositions.

XX Claim 1; Page; 25pp; English.

XX This invention describes novel Bacillus sp. alkaline proteases useful in detergent compositions, especially in laundry, bleaching or automatic dishwasher detergents. The novel proteases have an increased detergency (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This sequence represents a variant of the alkaline protease KP43 from Bacillus sp strain KSM-KP43 created from the wild-type protease represented in AAM50080 NOTE: This sequence is not represented in the specification but has been constructed from the sequence represented in record AAM50080

XX Sequence 434 AA;

Query Match 92.2%; Score 2071; DB 5; Length 434;
Best Local Similarity 94.0%; Pred. No. 5.2e-146;
Matches 408; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSVGLYGQGIIVAVADTGLDTRNDSSMHEAPRGKITALYALGRTN 60
DB 1 NDVARGIVKADVAQSSVGLYGQGIIVAVADTGLDTRNDSSMHEAPRGKITALYALGRTN 60

QY 61 NANDTNGHGHVAGSVLNGSTNKGMAQANLVFQISIMDSGGGLGSLPSNLQTLFSQAYS 120
DB 61 NANDTNGHGHVAGSVLNGSTNKGMAQANLVFQISIMDSGGGLGSLPSNLQTLFSQAYS 120

QY 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180

QY 181 TVGATENLRPFGSYADNINHVAFQSSRGPTKDGRIKPDVMAFGTIFLSARSSLAPDSSF 240
DB 181 TVGATENLRPFGSYADNINHVAFQSSRGPTKDGRIKPDVMAFGTIFLSARSSLAPDSSF 240

QY 241 WANHDSKIYAMGTSMTATPIVAGVAQLREHFVKNRGITPKPSLLKAALIAGNADIGLY 300
DB 241 WANHDSKIYAMGTSMTATPIVAGVAQLREHFVKNRGITPKPSLLKAALIAGNADIGLY 300

QY 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTQKATYSFTATAGKPLKISLVMSDAPASTTA 360
DB 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTQKATYSFTATAGKPLKISLVMSDAPASTTA 360

QY 361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNWDGNNVFNFINAPOSQGTITIEVQAYN 420
DB 361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNWDGNNVFNFINAPOSQGTITIEVQAYN 420

QY 421 VPVGPQTFSLAIVN 434
DB 421 VPVGPQTFSLAIVN 434

RESULT 12
AAM50084
ID AAM50084 standard; protein; 433 AA.

XX AC AAM50084;

XX DT 12-AUG-2002 (first entry)

XX DE Bacillus sp SD-521 (FERM BP-11162) alkaline protease protein fragment.

XX KW Alkaline protease; detergent; laundry; bleaching; dishwasher.

XX OS Bacillus sp.

XX FN EP1209233-A2.

XX PD 29-MAY-2002.

XX 22-NOV-2001; 2001BP-00127851.

XX 22-NOV-2000; 2000JP-00355166.

PR 12-APR-2001; 2001JP-00114048.

XX

PA (KAOS) KAO CORP.

XX Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;

PI Okuda M, Saeki K;

XX WPI; 2002-437518/47.

XX New modified alkaline proteases useful in detergent compositions.

XX Claim 5; Page 16-18; 25pp; English.

XX This invention describes novel Bacillus sp. alkaline proteases useful in

CC detergent compositions, especially in laundry, bleaching or automatic
CC dishwasher detergents. The novel proteases have an increased detergency &
CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
CC sequence represents a fragment of the alkaline protease SD-521 from
CC Bacillus sp strain SD-521 (FERM BP-11162) described in the method of the
CC invention
XX
XX
SQ Sequence 433 AA;

Query Match 88.9%; Score 1998.5; DB 5; Length 433;
Best Local Similarity 87.8%; Pred. No. 1.3e-140; Indels 1; Gaps 1;
Matches 381; Conservative 28; Mismatches 24;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRN 60
DB 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRN 60

QY 61 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NANTNGHGHVAGSVLNG - ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAWN 119

QY 121 AGARIHTNSGAAVNGAYTTDSRIVDVYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
DB 120 AGARIHTNSGAPVNGAYTANSRQVDEVVRNDMTVLFAAGNEGPNNGTISAPGTAKNAI 179

QY 181 TVGATENLRPSFGSYADNHNHVAQFSRGTGDKRIKPDVMAPTGTFILSARSLAPDSF 240
DB 180 TVGATENLRPSFGSLADNPNHIAQFSRGATRDGRKIPDVTAPGTTFILSARSLAPDSF 239

QY 241 WANHDSKYAVMGTSMTATPVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLY 300
DB 240 WANNYSKYAVMGTSMTATPVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLY 299

QY 301 PNGNQGWGRVTLKSLNVAIVYNESSLSSTOKATYSFTATAGKPLKISLVMSDAPASTTA 360
DB 300 PSGDQGWGRVTLKSLNVAIVYNEATALTGQKATYSFQAQKPLKISLVMTDAPGSTTA 359

QY 361 SVTLVNDLDLVTAPNGTQYVGNDFTSYNDNWDGRNNVENVFINAPQSGTYTIEVQAYN 420
DB 360 SYTLVNDLDLVTAPNGQKYVGNDFSYDNNWDGRNNVENVFINAPQSGTYTIEVQAYN 419

QY 421 VPVGQPTFSLAIVN 434
DB 420 VPSGPQRFSLAIVH 433

RESULT 13
ID AAM50082
AC AAM50082; protein; 433 AA.
XX AAM50082;
XX AAM50082;
DT 12-AUG-2002 (first entry)
XX Bacillus sp D6- (FERM P1592) alkaline protease protein fragment.
XX Alkaline protease; detergent; laundry; bleaching; dishwasher.
XX Bacillus sp.
XX EP1209233-A2.
XX 29-MAY-2002.
XX 22-NOV-2001; 2001EP-00127851.
XX 22-NOV-2000; 2000JP-00355166.
XX 12-APR-2001; 2001JP-00114048.
XX (KAOS) KAO CORP.
XX Harada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
PI Okuda M, Saeki K;

XX WPI; 2002-437518/47.
XX New modified alkaline proteases useful in detergent compositions.
XX Claim 5; Page 13-15; 25pp; English.
XX This invention describes novel Bacillus sp. alkaline proteases useful in
CC detergent compositions, especially in laundry, bleaching or automatic
CC dishwasher detergents. The novel proteases have an increased detergency &
CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
CC sequence represents a fragment of the alkaline protease E-1 from Bacillus
CC sp strain D6- (FERM-P1592) described in the method of the invention
XX
XX
SQ Sequence 433 AA;

Query Match 88.8%; Score 1994.5; DB 5; Length 433;
Best Local Similarity 87.6%; Pred. No. 2.6e-140; Indels 1; Gaps 1;
Matches 380; Conservative 28; Mismatches 25;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRN 60
DB 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRN 60

QY 61 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NANTNGHGHVAGSVLNG - ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAWN 119

QY 121 AGARIHTNSGAAVNGAYTTDSRIVDVYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
DB 120 AGARIHTNSGAPVNGAYTANSRQVDEVVRNDMTVLFAAGNEGPNNGTISAPGTAKNAI 179

QY 181 TVGATENLRPSFGSYADNHNHVAQFSRGTGDKRIKPDVMAPTGTFILSARSLAPDSF 240
DB 180 TVGATENLRPSFGSIADNPNHIAQFSRGATRDGRKIPDVTAPGTTFILSARSLAPDSF 239

QY 241 WANHDSKYAVMGTSMTATPVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLY 300
DB 240 WANNYSKYAVMGTSMTATPVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLY 299

QY 301 PNGNQGWGRVTLKSLNVAIVYNESSLSSTOKATYSFTATAGKPLKISLVMSDAPASTTA 360
DB 300 PSGDQGWGRVTLKSLNVAIVYNEATALTGQKATYSFQAQKPLKISLVMTDAPGSTTA 359

QY 361 SVTLVNDLDLVTAPNGTQYVGNDFTSYNDNWDGRNNVENVFINAPQSGTYTIEVQAYN 420
DB 360 SYTLVNDLDLVTAPNGQKYVGNDFSYDNNWDGRNNVENVFINAPQSGTYTIEVQAYN 419

QY 421 VPVGQPTFSLAIVN 434
DB 420 VPSGPQRFSLAIVH 433

RESULT 14
ID AAM89548
AC AAM89548; standard; protein; 636 AA.
XX AAM89548;
XX AAM89548;
DT 12-APR-1999 (first entry)
XX Bacillus sp. alkaline protease Y.
XX Alkaline protease Y; detergent; surfactant; leather processing;
KW debittering; flavour.
XX Bacillus sp.
XX WO9856927-A2.
XX 17-DEC-1998.
XX 09-JUN-1998; 98WO-US012005.
PF

XX 12-JUN-1997; 97US-00873479.
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 XX Sloma A, Christianson L;
 XX WPI; 1999-080908/07.
 XX Novel protease from *Bacillus subtilis* LC20 - useful in laundry and
 PT dishwashing detergents and for leather processing.
 XX Claim 3; Page 55-56; 77pp; English.
 XX This is the amino acid sequence of a *Bacillus* sp. alkaline protease Y
 CC that is said to have good alkali and surfactant resistance and improved
 CC detergency. It shows 77% identity to a newly isolated protease (see
 CC AA89547) of *Bacillus* sp. JPI70 (NCIB 12513). The invention provides
 CC vectors, recombinant host cells and methods for the recombinant
 CC production of such proteases. The protease are used in laundry and
 CC dishwashing detergents, for institutional and industrial cleaning, and
 CC for leather processing, as well as for debittering and enhancing the
 CC degree of hydrolysis of protein hydrolysates, for flavour development
 CC through hydrolysis of proteins, degradation of undesired peptides and in
 CC enzymatic synthesis of peptides. They have enhanced stability towards
 CC oxidation under alkaline conditions, e.g. towards bleaching agents of the
 CC peroxy type. The invention also provides mutant cells in which the
 CC protease activity is diminished. Such cells can be used for the
 CC production of heterologous recombinant proteins
 XX
 SQ Sequence 636 AA;
 Query Match 88.5%; Score 1989.5; DB 2; Length 636;
 Best Local Similarity 87.3%; Pred. No. 1e-139;
 Matches 379; Conservative 30; Mismatches 24; Indels 1; Gaps 1;
 QY 1 NDVARGIVKADVAQSSVGLVGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
 DB 204 NDVARGIVKADVAQNNYGLYGQQLVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 263
 QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLTLFSQAYS 120
 DB 264 NASDPNGHGHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGSLPSNLTLFSQAWN 322
 QY 121 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 323 AGARIHTNSGAPVNGAYTANSRQVDEYVRNNDMTILFAAGNEGPNSGTISAPGTAKNAI 382
 QY 181 TVGATENLRPSFGSYADNINNHVAFSSRGPTKDGRIKPDVMAPTGTFILSARSSLAPDSF 240
 DB 383 TVGATENYRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVTAPTGTFILSARSSLAPDSF 442
 QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 DB 443 WANYSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 502
 QY 301 PNGNQGWRTLDKSLNVAAYNNESSLSSTQATYSFTATAGKPLKISLVWSDAPASTTA 360
 DB 503 PNGDQGWRTLDKSLNVAAYNEATALATGQATYSFQAQAGKPLKISLVWTDAPGSTTA 562
 QY 361 SVTLVNDLVLITAPNGTQVGNDFSPYNDWGRNENVENFINAPQSGTYYIEVQAYN 420
 DB 563 SYTLVNDLVLITAPNGQYVGNDFSAFYDNNWGRNENVENFINAPQSGTYYIEVQAYN 622
 QY 421 VPVGPQTFSLAIYN 434
 DB 623 VPSGPQTFSLAIHV 636
 RESULT 15
 ID AAM50083
 XX AAM50083 standard; protein; 433 AA.

AC AAM50083;
 XX 12-AUG-2002 (first entry)
 XX *Bacillus* sp Y-(FERM BP-1029) alkaline protease protein fragment.
 DE Alkaline protease; detergent; laundry; bleaching; dishwasher.
 XX *Bacillus* sp.
 OS EPI209233-A2.
 XX 29-MAY-2002.
 PD 22-NOV-2001; 2001EP-00127851.
 PF 22-NOV-2000; 2000JP-00355166.
 PR 12-APR-2001; 2001JP-00114048.
 XX (KAOS) KAO CORP.
 PA Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
 PI Okuda M, Saeki K;
 XX WPI; 2002-437518/47.
 DR New modified alkaline proteases useful in detergent compositions.
 XX Claim 5; Page 15-16; 25pp; English.
 XX This invention describes novel *Bacillus* sp. alkaline proteases useful in
 CC detergent compositions, especially in laundry, bleaching or automatic
 CC dishwasher detergents. The novel proteases have an increased detergency &
 CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
 CC sequence represents a fragment of the alkaline protease Ya from *Bacillus*
 CC sp strain Y-(FERM BP-1029) described in the method of the invention
 XX
 SQ Sequence 433 AA;
 Query Match 88.5%; Score 1987.5; DB 5; Length 433;
 Best Local Similarity 87.3%; Pred. No. 8.7e-140;
 Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;
 QY 1 NDVARGIVKADVAQSSVGLVGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
 DB 1 NDVARGIVKADVAQNNYGLYGQQLVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
 QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLTLFSQAYS 120
 DB 61 NASDPNGHGHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGSLPSNLTLFSQAWN 119
 QY 121 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 120 AGARIHTNSGAPVNGAYTANSRQVDEYVRNNDMTILFAAGNEGPNSGTISAPGTAKNAI 179
 QY 181 TVGATENLRPSFGSYADNINNHVAFSSRGPTKDGRIKPDVMAPTGTFILSARSSLAPDSF 240
 DB 180 TVGATENYRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVTAPTGTFILSARSSLAPDSF 239
 QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 DB 240 WANYSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 299
 QY 301 PNGNQGWRTLDKSLNVAAYNNESSLSSTQATYSFTATAGKPLKISLVWSDAPASTTA 360
 DB 300 PNGDQGWRTLDKSLNVAAYNEATALATGQATYSFQAQAGKPLKISLVWTDAPGSTTA 359
 QY 361 SVTLVNDLVLITAPNGTQVGNDFSPYNDWGRNENVENFINAPQSGTYYIEVQAYN 420
 DB 360 SYTLVNDLVLITAPNGQYVGNDFSAFYDNNWGRNENVENFINAPQSGTYYIEVQAYN 419
 QY 421 VPVGPQTFSLAIYN 434
 DB 421 VPVGPQTFSLAIHV 434

Db 420 VPSGPORFSLAIVH 433

Search completed: March 10, 2004, 14:44:48
Job time : 61 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:43:43 ; Search time 22 Seconds
(without alignments)
1018.440 Million cell updates/sec

Title: US-09-985-689A-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVQAYNVFVGPQTFSLAIVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2247	100.0	640	4	US-09-509-814A-6
2	2242	99.8	640	4	US-09-509-814A-8
3	2183	97.2	639	4	US-09-509-814A-4
4	2155	95.9	639	4	US-09-509-814A-1
5	2155	95.9	640	4	US-09-509-814A-2
6	2125.5	94.6	641	2	US-08-873-479-42
7	1986.5	88.4	433	4	US-09-104-623A-4
8	1986.5	88.4	433	4	US-09-019-532-4
9	1986.5	88.4	433	4	US-09-338-746-4
10	1986.5	88.4	635	2	US-08-873-479-43
11	1581.5	70.4	345	4	US-09-512-251A-10
12	1581.5	70.4	345	4	US-09-515-150A-10
13	1581.5	70.4	345	4	US-09-196-281-13
14	452.5	20.1	659	3	US-08-894-818B-1
15	452.5	20.1	659	4	US-09-445-472-12
16	414	18.4	412	4	US-09-445-472-1
17	414	18.4	522	3	US-08-894-818B-3
18	414	18.4	522	4	US-09-445-472-4
19	414	18.4	654	3	US-08-894-818B-35
20	414	18.4	654	4	US-09-445-472-16
21	401	17.8	659	3	US-08-894-818B-5
22	346	15.4	520	3	US-09-000-016-7
23	346	15.4	520	4	US-09-514-340-7
24	346	15.4	734	3	US-09-000-016-4
25	346	15.4	734	4	US-09-514-340-4
26	346	15.4	823	3	US-09-000-016-2
27	346	15.4	823	4	US-09-514-340-2

28	304.5	13.6	903	1	US-08-750-532-1	Sequence 1, Appli
29	304.5	13.6	1398	1	US-08-750-532-9	Sequence 9, Appli
30	304.5	13.6	1398	3	US-08-894-818B-8	Sequence 8, Appli
31	304.5	13.6	1398	4	US-09-445-472-6	Sequence 6, Appli
32	295.5	13.2	237	1	US-08-750-532-18	Sequence 18, Appli
33	283.5	12.6	418	2	US-08-873-479-44	Sequence 44, Appli
34	276	12.3	418	4	US-09-966-921A-2	Sequence 2, Appli
35	256.5	11.4	397	4	US-09-328-352-7533	Sequence 7533, Ap
36	254.5	11.3	275	1	US-08-431-387-1	Sequence 1, Appli
37	251	11.2	280	1	US-08-434-255-8	Sequence 8, Appli
38	251	11.2	280	1	US-08-459-967-8	Sequence 8, Appli
39	251	11.2	280	1	US-08-460-327-8	Sequence 8, Appli
40	251	11.2	280	1	US-08-459-871-8	Sequence 8, Appli
41	251	11.2	280	3	US-09-024-532-2	Sequence 2, Appli
42	251	11.2	280	4	US-09-104-623A-2	Sequence 2, Appli
43	251	11.2	280	4	US-09-019-532-2	Sequence 2, Appli
44	251	11.2	280	4	US-09-417-359A-2	Sequence 2, Appli
45	251	11.2	280	4	US-09-705-185-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-509-814A-6
; Sequence 6, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUM
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-OPCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 640
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-509-814A-6

Query Match 100.0%; Score 2247; DB 4; Length 640;
Best Local Similarity 100.0%; Pred. No. 3.le-173;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	NDVARGIVKADVAQSSVGLYGGQIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN	60
Db	207	NDVARGIVKADVAQSSVGLYGGQIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN	266
Qy	61	NANDTNGHGHVAGSVLGNSTKGMAPQANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS	120
Db	267	NANDTNGHGHVAGSVLGNSTKGMAPQANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS	326
Qy	121	AGARIHTNSWGAANVAGYTTDSRNVDYVRKNDMTILFAAGNEGNGGTISAPGTAKNAI	180
Db	327	AGARIHTNSWGAANVAGYTTDSRNVDYVRKNDMTILFAAGNEGNGGTISAPGTAKNAI	386
Qy	181	TVGATENTLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMPAGTIFILSARSSLPDSSF	240
Db	387	TVGATENTLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMPAGTIFILSARSSLPDSSF	446
Qy	241	WANHDSKYAIMGTSWATPIVAGNVAQLREHFVKRGITPKPSLLKAALIAGAADIGLGY	300

Db 447 WANHDSKYAMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAAL1AGAADIGLGY 506
QY 301 PNGNQGWRTLDKSLNVAAYNSSLTSQKATYSFTATAGPKPLKISLWSDAPASTTA 360
Db 507 PNGNQGWRTLDKSLNVAAYNSSLTSQKATYSFTATAGPKPLKISLWSDAPASTTA 566
QY 361 SVTLVNDLVLITAPNGTQYVGNDFTSFYNDNDGRNENVENFINAPQSGTYTIEVQAYN 420
Db 567 SVTLVNDLVLITAPNGTQYVGNDFTSFYNDNDGRNENVENFINAPQSGTYTIEVQAYN 626
QY 421 VPVGPQTFSLAIVN 434
Db 627 VPVGPQTFSLAIVN 640

RESULT 2

US-09-509-814A-8

; Sequence 8, Application US/09509814A

; Patent No. 6376227

; GENERAL INFORMATION:

; APPLICANT: TAKAIWA, MIKIO

; APPLICANT: OKUDA, MITSUYOSHI

; APPLICANT: SAEKI, KATSUHISA

; APPLICANT: KUBOTA, HIROMI

; APPLICANT: HITOMI, JUN

; APPLICANT: KAGEYAMA, YASUSHI

; APPLICANT: SHIKATA, SHITSUMI

; APPLICANT: NOMURA, MASAFUMI

; TITLE OF INVENTION: ALKALINE PROTEASE

; FILE REFERENCE: 0327-0832-0PCT

; CURRENT APPLICATION NUMBER: US/09/509,814A

; CURRENT FILING DATE: 2000-04-06

; PRIOR APPLICATION NUMBER: PCT/JP98/04528

; PRIOR FILING DATE: 1998-10-07

; PRIOR APPLICATION NUMBER: JP 9-274570

; PRIOR FILING DATE: 1997-06-08

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 8

; LENGTH: 640

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-509-814A-8

Query Match 99.8%; Score 2242; DB 4; Length 640;

Best Local Similarity 99.8%; Pred. No. 7.8e-173; Mismatches 1; Indels 0; Gaps 0;

Matches 433; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAPRGKITAYALGRTN 60
Db 207 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAPRGKITAYALGRTN 266
QY 61 NANTNGHGHVAGSVLNGSTNKGAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 267 NANTNGHGHVAGSVLNGSTNKGAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 326
QY 121 AGARIHTNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 327 AGARIHTNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTILSARSLAPDSF 240
Db 387 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTILSARSLAPDSF 446
QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAAL1AGAADIGLGY 300
Db 447 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAAL1AGAADIGLGY 506
QY 301 PNGNQGWRTLDKSLNVAAYNSSLTSQKATYSFTATAGPKPLKISLWSDAPASTTA 360
Db 507 PNGNQGWRTLDKSLNVAAYNSSLTSQKATYSFTATAGPKPLKISLWSDAPASTTA 566

QY 361 SVTLVNDLVLITAPNGTQYVGNDFTSFYNDNDGRNENVENFINAPQSGTYTIEVQAYN 420
Db 567 SVTLVNDLVLITAPNGTQYVGNDFTSFYNDNDGRNENVENFINAPQSGTYTIEVQAYN 626
QY 421 VPVGPQTFSLAIVN 434
Db 627 VPVGPQTFSLAIVN 640

RESULT 3

US-09-509-814A-4

; Sequence 4, Application US/09509814A

; Patent No. 6376227

; GENERAL INFORMATION:

; APPLICANT: TAKAIWA, MIKIO

; APPLICANT: OKUDA, MITSUYOSHI

; APPLICANT: SAEKI, KATSUHISA

; APPLICANT: KUBOTA, HIROMI

; APPLICANT: HITOMI, JUN

; APPLICANT: KAGEYAMA, YASUSHI

; APPLICANT: SHIKATA, SHITSUMI

; APPLICANT: NOMURA, MASAFUMI

; TITLE OF INVENTION: ALKALINE PROTEASE

; FILE REFERENCE: 0327-0832-0PCT

; CURRENT APPLICATION NUMBER: US/09/509,814A

; CURRENT FILING DATE: 2000-04-06

; PRIOR APPLICATION NUMBER: PCT/JP98/04528

; PRIOR FILING DATE: 1998-10-07

; PRIOR APPLICATION NUMBER: JP 9-274570

; PRIOR FILING DATE: 1997-06-08

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 4

; LENGTH: 639

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-509-814A-4

Query Match 97.2%; Score 2183; DB 4; Length 639;

Best Local Similarity 96.3%; Pred. No. 4.5e-168; Mismatches 3; Indels 0; Gaps 0;

Matches 418; Conservative 13; Mismatches 3; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAPRGKITAYALGRTN 60
Db 206 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAPRGKITAYALGRTN 265
QY 61 NANTNGHGHVAGSVLNGSTNKGAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 266 NANTNGHGHVAGSVLNGSTNKGAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 325
QY 121 AGARIHTNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 326 AGARIHTNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 385
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTILSARSLAPDSF 240
Db 386 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTILSARSLAPDSF 445
QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAAL1AGAADIGLGY 300
Db 446 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAAL1AGAADIGLGY 505
QY 301 PNGNQGWRTLDKSLNVAAYNSSLTSQKATYSFTATAGPKPLKISLWSDAPASTTA 360
Db 506 PNGNQGWRTLDKSLNVAAYNSSLTSQKATYSFTATAGPKPLKISLWSDAPASTTA 565
QY 361 SVTLVNDLVLITAPNGTQYVGNDFTSFYNDNDGRNENVENFINAPQSGTYTIEVQAYN 420
Db 566 SVTLVNDLVLITAPNGTQYVGNDFTSFYNDNDGRNENVENFINAPQSGTYTIEVQAYN 625
QY 421 VPVGPQTFSLAIVN 434
Db 626 VPVGPQTFSLAIVN 639

RESULT 4
US-09-509-814A-1
Sequence 1, Application US/09509814A
Patent No. 6376227
GENERAL INFORMATION:
APPLICANT: TAKAIWA, MIKIO
APPLICANT: OKUDA, MITSUYOSHI
APPLICANT: SAEKI, KATSUHISA
APPLICANT: KUBOTA, HIROMI
APPLICANT: HITOMI, JUN
APPLICANT: KAGEYAMA, YASUSHI
APPLICANT: SHIKATA, SHITSUMI
APPLICANT: NOMURA, MASAFUMI
TITLE OF INVENTION: ALKALINE PROTEASE
FILE REFERENCE: 0327-0832-0PCT
CURRENT APPLICATION NUMBER: US/09/509, 814A
CURRENT FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/JP98/04528
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: JP 9-274570
PRIOR FILING DATE: 1997-06-08
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 639
TYPE: PRT
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (23)..(23)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (29)..(29)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (32)..(32)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (46)..(46)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (47)..(47)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (53)..(53)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (70)..(70)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (74)..(74)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (89)..(89)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (102)..(102)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (105)..(105)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (128)..(128)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (130)..(130)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (131)..(131)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (132)..(132)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (133)..(133)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (146)..(146)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (148)..(148)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (160)..(160)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (165)..(165)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (172)..(172)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (183)..(183)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (187)..(187)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (188)..(188)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (189)..(189)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (194)..(194)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (286)..(286)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (306)..(306)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (324)..(324)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (369)..(369)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (431)..(431)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (501)..(501)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (531)..(531)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (541)..(541)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (584)..(584)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (591)..(591)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (592)..(592)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (594)..(594)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (595)..(595)

OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (396)..(596)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (611)..(611)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (632)..(632)
OTHER INFORMATION: Xaa is any amino acid
US-09-509-814A-1

Query Match 95.9%; Score 2155; DE 4; Length 639;
Best Local Similarity 96.3%; Pred. No. 8.3e-166;
Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 1 NDVARGIVKADVAQSSYGLYGQIVAVADTGLDGRNDSSHEAFRGKITALVALGRTN 60
Db 206 NDVARGIVKADVAQSSYGLYGQIVAVADTGLDGRNDSSHEAFRGKITALVALGRTN 265
QY 61 NANDNGHGHVAGSVLGNSTNGKGMPOANLVFOSIMDSGGLGGLPSNLQTLFQAYS 120
Db 266 NANDNGHGHVAGSVLGNSTNGKGMPOANLVFOSIMDSGGLGGLPSNLQTLFQAYS 325
QY 121 AGARIHTNSWGAANGAYTTDSRNVDDYVRKDMTILFAAGNEGNGGTISAPGTAKNAI 180
Db 326 AGARIHTNSWGAANGAYTTDSRNVDDYVRKDMTILFAAGNEGNGGTISAPGTAKNAI 385
QY 181 TVGATENLEPSPGSGVADNHNHVAQFSSRGPTKDGRIKPDVMAPTGFIILSARSSSLAPDSFF 240
Db 386 TVGATENLEPSPGSGVADNHNHVAQFSSRGPTKDGRIKPDVMAPTGFIILSARSSSLAPDSFF 445
QY 241 WANHSKYAIMGSTWATPIVAGNVAQUREHFVKNRGITPKPSLLKAALIRAGAADIGLY 300
Db 446 WANHSKYAIMGSTWATPIVAGNVAQUREHFVKNRGITPKPSLLKAALIRAGAADIGLY 505
QY 301 PNGNQGWGRTLDKSLNVAYNVESLSSTQKATYSFTATAGPKIKISLVWSDAPASTTA 360
Db 506 PNGNQGWGRTLDKSLNVAYNVESLSSTQKATYSFTATAGPKIKISLVWSDAPASTTA 565
QY 361 SVTLVNDLVLITAPNGTYGVNDFTSPYNDWNGRNNVNFVINAPOSQGTITIEVQAYN 420
Db 566 SVTLVNDLVLITAPNGTYGVNDFTSPYNDWNGRNNVNFVINAPOSQGTITIEVQAYN 625
QY 421 VPVGQPTFSLAVN 434
Db 626 VPVGQPTFSLAVN 639

RESULT 5
US-09-509-814A-2
Sequence 2, Application US/09509814A
Patent No. 6376227
GENERAL INFORMATION:
APPLICANT: OKUDA, MIKIO
APPLICANT: OKUDA, MITSUYOSHI
APPLICANT: SAEKI, KATSUHIKA
APPLICANT: KUBOTA, HIROMI
APPLICANT: HITOMI, JUN
APPLICANT: KAGEYAMA, YASUSHI
APPLICANT: SHIKATA, SHITSUW
APPLICANT: NOMURA, MASAFUMI
TITLE OF INVENTION: ALKALINE PROTEASE
FILE REFERENCE: 0327-0832-OPCT
CURRENT APPLICATION NUMBER: US/09/509,814A
CURRENT FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/JP98/04528
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: JP 9-274570
PRIOR FILING DATE: 1997-06-08
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2

LENGTH: 640
TYPE: PRT
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: misc feature
LOCATION: (3)..(3)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (24)..(24)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (30)..(30)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (33)..(33)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (47)..(47)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (48)..(48)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (54)..(54)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (71)..(71)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (75)..(75)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (90)..(90)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (103)..(103)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (106)..(106)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (129)..(129)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (131)..(131)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (132)..(132)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (133)..(133)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (134)..(134)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (147)..(147)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (149)..(149)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (161)..(161)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (166)..(166)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (173)..(173)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (184)..(184)
OTHER INFORMATION: Xaa is any amino acid

NAME/KEY: misc feature
LOCATION: (198)..(198)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (189)..(189)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (190)..(190)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (195)..(195)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (287)..(287)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (307)..(307)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (325)..(325)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (502)..(502)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (532)..(532)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (542)..(542)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (585)..(585)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (592)..(592)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (593)..(593)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (595)..(595)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (596)..(596)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (597)..(597)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (612)..(612)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (633)..(633)
OTHER INFORMATION: Xaa is any amino acid
US-09-509-814A-2

Query Match 95.9%; Score 2155; DB 4; Length 640;
Best Local Similarity 96.3%; Pred. No. 8.3e-166;
Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 NDVARGIKADVAGSSYGLYGGQIVAVDTGLTGRNDSMHFAFRGKITALYALGRTN 60
DB 207 NDVARGIKADVAGSSYGLYGGQIVAVDTGLTGRNDSMHFAFRGKITALYALGRTN 266

QY 61 NANDTNGHGTHTVAGSVLNGSTNGKMAPQANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
DB 267 NANDTNGHGTHTVAGSVLNGSTNGKMAPQANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 326

QY 121 AGARIHTNSWGAAVNGAYTTDSRNVDVYRNDMTILPAAGNEGPNGGTISAPGTAKNAI 180
DB 327 AGARIHTNSWGAAVNGAYTTDSRNVDVYRNDMTILPAAGNEGPNGGTISAPGTAKNAI 386
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVWAPGTFFILSARSSLPDSSF 240
DB 387 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVWAPGTFFILSARSSLPDSSF 446
QY 241 WANHDSKYAYMGTTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALIAGAADIGLGY 300
DB 447 WANHDSKYAYMGTTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALIAGAADIGLGY 506
QY 301 PNGNGWGRVTLDSKLNVAIVYNESSLSSTQKATYSFTATAGKPLKISLVWSDAPASTTA 360
DB 507 PNGNGWGRVTLDSKLNVAIVYNESSLSSTQKATYSFTATAGKPLKISLVWSDAPASTTA 566
QY 361 SVTLVNDLDLVTAPNGTQYVGNDFTSYNDNDWGRNNVENVFAPQSGTYTIEVQAYN 420
DB 567 SVTLVNDLDLVTAPNGTQYVGNDFTSYNDNDWGRNNVENVFAPQSGTYTIEVQAYN 626
QY 421 VVPGPQTFSLAIVN 434
DB 627 VVPGPQTFSLAIVN 640

RESULT 6
US-08-873-479-42
Sequence 42, Application US/08873479
Patent No. 5891701
GENERAL INFORMATION:
APPLICANT: Sloma, Alan
APPLICANT: Lyne, Christianson
TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
TITLE OF INVENTION: Having Protease Activity
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5891701 No. 5891701disk of No. 5891701th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/873,479
FILING DATE: 12-JUN-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Agrie, Cheryl H
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 5251.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 641 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-873-479-42

Query Match 94.6%; Score 2125.5; DB 2; Length 641;
Best Local Similarity 93.5%; Pred. No. 2e-163;
Matches 406; Conservative 19; Mismatches 8; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTN 60
Db 209 NDVARGIVKADVAQNFGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTN 268
QY 61 NANDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
Db 269 NANDTNGHGHVAGSVLNG-ATNKGWAPQANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 327
QY 121 AGARIHTNSWGAAVNGAYTTDSRNVDYVYRKNMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 328 AGARIHTNSWGAPVNGAYTTDSRNVDYVYRKNMTILFAAGNEGPNGGTISAPGTAKNAI 387
QY 181 TVGATENLRSPSGSYADNINHVAQFSSRGPTKDGRIKPDVMAQPTILSARSSSLAPDSSF 240
Db 388 TVGATENLRSPSGSYADNINHVAQFSSRGPTKDGRIKPDVMAQPTILSARSSSLAPDSSF 447
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY 300
Db 448 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY 507
QY 301 PNGNQGWGRVTLDKSLNVAVNNESSLSSTQKATYSTFTATAGPKPLKISLVWSDAPASTTA 360
Db 508 PNGNQGWGRVTLDKSLNVAVNNESSLSSTQKATYSTFTATAGPKPLKISLVWSDAPASTTA 567
QY 361 SVTLVNDLDELIVITAPNGTQVGNDFTSFYNDNDGRNNVNFVINAPOSCTYTIIEVOAYN 420
Db 568 SVTLVNDLDELIVITAPNGTQVGNDFTSFYNDNDGRNNVNFVINAPOSCTYTIIEVOAYN 627
QY 421 VPVGPOTFSLAIVN 434
Db 628 VPVGPOTFSLAIVH 641

RESULT 7

US-09-104-623A-4
; Sequence 4, Application US/09104623A
; Patent No. 6303752
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne Agerlin
; APPLICANT: Fatum, Tine Muxoll
; APPLICANT: Deussen, Heinz-Josef
; APPLICANT: Rosgen, Erwin Ludo
; TITLE OF INVENTION: A Modified Polypeptide
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6303752o No. 6303752disk of No. 6303752th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,623A
; FILING DATE: 25-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 5256.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; STRANDEDNESS:

TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; STRAIN: Bacillus sp. Y
US-09-104-623A-4
Query Match 88.4%; Score 1986.5; DB 4; Length 433;
Best Local Similarity 87.3%; Pred. No. 1.9e-152;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;
QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTN 60
Db 1 NDVARGIVKADVAQNNYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
Db 61 NASDPNGHGHVAGSVLNG-ALNKGWAPQANLVFQSIMDSGGGLGGLPSNLQTLFQAYN 119
QY 121 AGARIHTNSWGAAVNGAYTTDSRNVDYVYRKNMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 120 AGARIHTNSWGAPVNGAYTANSRQVDEYVRNNDMTVLFAAGNEGPNSGTISAPGTAKNAI 179
QY 181 TVGATENLRSPSGSYADNINHVAQFSSRGPTKDGRIKPDVMAQPTILSARSSSLAPDSSF 240
Db 180 TVGATENLRSPSGSYADNINHVAQFSSRGPTKDGRIKPDVMAQPTILSARSSSLAPDSSF 239
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY 300
Db 240 WANVNSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY 299
QY 301 PNGNQGWGRVTLDKSLNVAVNNESSLSSTQKATYSTFTATAGPKPLKISLVWSDAPASTTA 360
Db 300 PSGDQGWGRVTLDKSLNVAVNEATLQCKATYSTFQAQAGKPLKISLVWSDAPASTTA 359
QY 361 SVTLVNDLDELIVITAPNGTQVGNDFTSFYNDNDGRNNVNFVINAPOSCTYTIIEVOAYN 420
Db 360 SVTLVNDLDELIVITAPNGKTVGNDFSYPDNDNDGRNNVNFVINAPOSCTYTIIEVOAYN 419
QY 421 VPVGPOTFSLAIVN 434
Db 420 VPVGPOTFSLAIVH 433

RESULT 8

US-09-019-532-4
; Sequence 4, Application US/09019532B
; Patent No. 6416756
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne Agerlin
; APPLICANT: Prent, Annette
; TITLE OF INVENTION: A Modified Enzyme for Skin Care
; FILE REFERENCE: 4922.204-US
; CURRENT APPLICATION NUMBER: US/09/019,532B
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: 0038/97
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: 0754/97
; EARLIER FILING DATE: 1997-06-25
; EARLIER APPLICATION NUMBER: 60/051,381
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: PCT/DK98/00015
; EARLIER FILING DATE: 1998-01-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-019-532-4

Query Match 88.4%; Score 1986.5; DB 4; Length 433;
Best Local Similarity 87.3%; Pred. No. 1.9e-152;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

```
QY 1 NDVARGIVKADVAGSSVGLYGGQGVAVADTGLDGTGRDSSMHEAFRGKITALYALGRTN 60
Db 1 NDVARGIVKADVAGSSVGLYGGQGVAVADTGLDGTGRDSSMHEAFRGKITALYALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNGKMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
Db 61 NASDPNGHGHVAGSVLNGSTNGKMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 119
QY 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 120 AGARIHTNSGAPVNGAYTANSRQVDEYVRNDMTILFAAGNEGPNGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADININHVAFSSRGPTKDGRIKPDVMAFGTIFILSARSLAPDSSF 240
Db 180 TVGATENLRPSFGSIADPNPHIAQFSSRGATRDGRIKPDVTAFGTIFILSARSLAPDSSF 239
QY 241 WANHSKYAYMGTSMATPIVAGNVAQLREHFVKRGTITPKPSLLKAALIAGAADIGLY 300
Db 240 WANTNSKYAYMGTSMATPIVAGNVAQLREHFVKRGTITPKPSLLKAALIAGAADIGLY 299
QY 301 PNGNGGGRVTLDKSLNVAYNVNESSLSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
Db 300 PSGDQGGGRVTLDKSLNVAYNVNEATATGOKATYSFOAQAGKPLKISLVWTDAPGSTTA 359
QY 361 SYTLVNDLVLITAPNGTQYVGNDFSTPYDNDNDGRNNVNFINAPQSGTYTIEVQAYN 420
Db 360 SYTLVNDLVLITAPNGQKYVGNDFSYPDNDNDGRNNVNFINAPQSGTYTIEVQAYN 419
QY 421 VPVGPQTFSLAIVN 434
Db 420 VPSGQRFSLAIVH 433

RESULT 9
US-09-338-746-4
; Sequence 4, Application US/09338746
; Patent No. 6638526
; GENERAL INFORMATION:
; APPLICANT: Deussen, Heinz-Josef
; APPLICANT: Olsen, Arne A.
; APPLICANT: Fatum, Tine M.
; APPLICANT: Roggen, Erwin L.
; TITLE OF INVENTION: A Polypeptide-Polymer Conjugate
; FILE REFERENCE: 5619,200-US
; CURRENT APPLICATION NUMBER: US/09/338,746
; EARLIER FILING DATE: 1999-06-23
; EARLIER APPLICATION NUMBER: PA 1998 00809
; EARLIER FILING DATE: 1998-06-23
; EARLIER APPLICATION NUMBER: 60/091,461
; EARLIER FILING DATE: 1998-07-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-338-746-4

Query Match 88.4%; Score 1986.5; DB 4; Length 433;
Best Local Similarity 87.3%; Pred. No. 1.9e-152;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAGSSVGLYGGQGVAVADTGLDGTGRDSSMHEAFRGKITALYALGRTN 60
Db 1 NDVARGIVKADVAGSSVGLYGGQGVAVADTGLDGTGRDSSMHEAFRGKITALYALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNGKMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
Db 61 NASDPNGHGHVAGSVLNGSTNGKMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 119
QY 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 120 AGARIHTNSGAPVNGAYTANSRQVDEYVRNDMTILFAAGNEGPNGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADININHVAFSSRGPTKDGRIKPDVMAFGTIFILSARSLAPDSSF 240
Db 180 TVGATENLRPSFGSIADPNPHIAQFSSRGATRDGRIKPDVTAFGTIFILSARSLAPDSSF 239
QY 241 WANHSKYAYMGTSMATPIVAGNVAQLREHFVKRGTITPKPSLLKAALIAGAADIGLY 300
Db 240 WANTNSKYAYMGTSMATPIVAGNVAQLREHFVKRGTITPKPSLLKAALIAGAADIGLY 299
QY 301 PNGNGGGRVTLDKSLNVAYNVNESSLSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
Db 300 PSGDQGGGRVTLDKSLNVAYNVNEATATGOKATYSFOAQAGKPLKISLVWTDAPGSTTA 359
QY 361 SYTLVNDLVLITAPNGTQYVGNDFSTPYDNDNDGRNNVNFINAPQSGTYTIEVQAYN 420
Db 360 SYTLVNDLVLITAPNGQKYVGNDFSYPDNDNDGRNNVNFINAPQSGTYTIEVQAYN 419
QY 421 VPVGPQTFSLAIVN 434
Db 420 VPSGQRFSLAIVH 433
```

```
Db 120 AGARIHTNSGAPVNGAYTANSRQVDEYVRNDMTILFAAGNEGPNSGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADININHVAFSSRGPTKDGRIKPDVMAFGTIFILSARSLAPDSSF 240
Db 180 TVGATENLRPSFGSIADPNPHIAQFSSRGATRDGRIKPDVTAFGTIFILSARSLAPDSSF 239
QY 241 WANHSKYAYMGTSMATPIVAGNVAQLREHFVKRGTITPKPSLLKAALIAGAADIGLY 300
Db 240 WANTNSKYAYMGTSMATPIVAGNVAQLREHFVKRGTITPKPSLLKAALIAGAADIGLY 299
QY 301 PNGNGGGRVTLDKSLNVAYNVNESSLSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
Db 300 PSGDQGGGRVTLDKSLNVAYNVNEATATGOKATYSFOAQAGKPLKISLVWTDAPGSTTA 359
QY 361 SYTLVNDLVLITAPNGTQYVGNDFSTPYDNDNDGRNNVNFINAPQSGTYTIEVQAYN 420
Db 360 SYTLVNDLVLITAPNGQKYVGNDFSYPDNDNDGRNNVNFINAPQSGTYTIEVQAYN 419
QY 421 VPVGPQTFSLAIVN 434
Db 420 VPSGQRFSLAIVH 433

RESULT 10
US-08-873-479-43
; Sequence 43, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; APPLICANT: Lytne, Christianson
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5891701o No. 5891701disk of No. 5891701th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/873,479
; FILING DATE: 12-JUN-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Agrie, Cheryl H
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 5251,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-873-479-43

Query Match 88.4%; Score 1986.5; DB 2; Length 635;
Best Local Similarity 87.3%; Pred. No. 3.3e-152;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAGSSVGLYGGQGVAVADTGLDGTGRDSSMHEAFRGKITALYALGRTN 60
Db 203 NDVARGIVKADVAGSSVGLYGGQGVAVADTGLDGTGRDSSMHEAFRGKITALYALGRTN 262
```

```
QY 61 NNDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLPSNLQTLFSQAYS 120
Db 263 NADDFNGHGHVAGSVLNG-ALNKGWAPQANLVFQSIMDSGGGLPSNLQTLFSQAWN 321
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 322 AGARIHTNSGAAVNGAYTANSRQVDEYVRNNDMTILFAAGNEGPNGGTISAPGTAKNAI 381
QY 181 TVGATENLPSFGSYADNHNHVAQFSSRGTGDKRIKPDVMAFGTIFLSARSLLAPDSSF 240
Db 382 TVGATENLPSFGSYADNHNHVAQFSSRGTGDKRIKPDVMAFGTIFLSARSLLAPDSSF 441
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
Db 442 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 501
QY 301 PNGNQGWGRVTLDKSLNVAYNESLSSTOKATYSFTATAGKPLKISLVWSDAASITA 360
Db 502 PNGNQGWGRVTLDKSLNVAYNESLSSTOKATYSFTATAGKPLKISLVWSDAASITA 561
QY 361 SVTLVNDLVLITAPNGTQVGNDFTPYNDNWDGNNVNFVINAPOSQGTITIEVQAYN 420
Db 562 SVTLVNDLVLITAPNGKQVGNDFPYNDNWDGNNVNFVINAPOSQGTITIEVQAYN 621
QY 421 VFGVQPTFSLAIYN 434
Db 622 VPSGQRFSLAIYH 635

RESULT 11
US-09-512-251A-10
; Sequence 10, Application US/09512251A
; Patent No. 655355
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; APPLICANT: Andersen, Kim
; TITLE OF INVENTION: Protease Variants and Compositions
; FILE REFERENCE: 5349.204-US
; CURRENT APPLICATION NUMBER: US/09/512.251A
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-09-512-251A-10

Query Match 70.4%; Score 1581.5; DB 4; Length 345;
Best Local Similarity 95.3%; Pred. No. 7.1e-120;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
Db 29 NDVARGIVKADVAQNNFGLYGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 88
QY 61 NNDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLPSNLQTLFSQAYS 120
Db 89 NNDTNGHGHVAGSVLNG-ATNKGWAPQANLVFQSIMDSGGGLPSNLQTLFSQAYS 147
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 148 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 207
QY 181 TVGATENLPSFGSYADNHNHVAQFSSRGTGDKRIKPDVMAFGTIFLSARSLLAPDSSF 240
Db 208 TVGATENLPSFGSYADNHNHVAQFSSRGTGDKRIKPDVMAFGTIFLSARSLLAPDSSF 267
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
Db 268 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 327

RESULT 13
US-09-136-281-13
; Sequence 13, Application US/09196281A
; Patent No. 6605458
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter K.
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; TITLE OF INVENTION: Protease Variants And Compositions
; FILE REFERENCE: 5435.200-US
; CURRENT APPLICATION NUMBER: US/09/196.281A
; CURRENT FILING DATE: 1998-11-19
; EARLIER APPLICATION NUMBER: 1332/97
; EARLIER FILING DATE: 1997-11-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
```

```
QY 61 NNDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLPSNLQTLFSQAYS 120
Db 263 NADDFNGHGHVAGSVLNG-ALNKGWAPQANLVFQSIMDSGGGLPSNLQTLFSQAWN 321
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 322 AGARIHTNSGAAVNGAYTANSRQVDEYVRNNDMTILFAAGNEGPNGGTISAPGTAKNAI 381
QY 181 TVGATENLPSFGSYADNHNHVAQFSSRGTGDKRIKPDVMAFGTIFLSARSLLAPDSSF 240
Db 382 TVGATENLPSFGSYADNHNHVAQFSSRGTGDKRIKPDVMAFGTIFLSARSLLAPDSSF 441
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
Db 442 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 501
QY 301 PNGNQGWGRVTLDKSLNVAYNESLSSTOKATYSFTATAGKPLKISLVWSDAASITA 360
Db 502 PNGNQGWGRVTLDKSLNVAYNESLSSTOKATYSFTATAGKPLKISLVWSDAASITA 561
QY 361 SVTLVNDLVLITAPNGTQVGNDFTPYNDNWDGNNVNFVINAPOSQGTITIEVQAYN 420
Db 562 SVTLVNDLVLITAPNGKQVGNDFPYNDNWDGNNVNFVINAPOSQGTITIEVQAYN 621
QY 421 VFGVQPTFSLAIYN 434
Db 622 VPSGQRFSLAIYH 635

RESULT 11
US-09-512-251A-10
; Sequence 10, Application US/09512251A
; Patent No. 655355
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; APPLICANT: Andersen, Kim
; TITLE OF INVENTION: Protease Variants and Compositions
; FILE REFERENCE: 5349.204-US
; CURRENT APPLICATION NUMBER: US/09/512.251A
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-09-512-251A-10

Query Match 70.4%; Score 1581.5; DB 4; Length 345;
Best Local Similarity 95.3%; Pred. No. 7.1e-120;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
Db 29 NDVARGIVKADVAQNNFGLYGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 88
QY 61 NNDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLPSNLQTLFSQAYS 120
Db 89 NNDTNGHGHVAGSVLNG-ATNKGWAPQANLVFQSIMDSGGGLPSNLQTLFSQAYS 147
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 148 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 207
QY 181 TVGATENLPSFGSYADNHNHVAQFSSRGTGDKRIKPDVMAFGTIFLSARSLLAPDSSF 240
Db 208 TVGATENLPSFGSYADNHNHVAQFSSRGTGDKRIKPDVMAFGTIFLSARSLLAPDSSF 267
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
Db 268 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 327
```

```
/ LENGTH: 345
/ TYPE: PRT
/ ORGANISM: Bacillus
US-09-196-281-13

Query Match      70.4%; Score 1581.5; DB 4; Length 345;
Best Local Similarity 95.3%; Pred. No. 7.1e-120;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSVGLYGQGIQVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 29 NDVARGIVKADVAQNNFGLYGQGIQVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 88
QY 61 NADNTNGHGHVAGSVLNGSTNKGAPQANLVFQSIMDSGGGLGLPSNLQTLFSQAYS 120
Db 89 NADNPNGHGHVAGSVLNG-ATNKGAPQANLVFQSIMDSGGGLGLPSNLQTLFSQAYS 147
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPGGTISAPGTAKNAI 180
Db 148 AGARIHTNSGAPVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPGGTISAPGTAKNAI 207
QY 181 TVGATENLRPSFGSYADNINHVAQFSRGTGDKRIKPDVMAFGTILSARSSLAPDSF 240
Db 208 TVGATENLRPSFGSYADNINHVAQFSRGTGDKRIKPDVMAFGTILSARSSLAPDSF 267
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADIGLGY 300
Db 268 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADVGLGF 327
QY 301 PNGNQGWGRVTLDKSLNV 318
Db 328 PNGNQGWGRVTLDKSLNV 345

RESULT 14
US-08-894-8183-1
; Sequence 1, Application US/088948183
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618

; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 659 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-894-818B-1

Query Match      20.1%; Score 452.5; DB 3; Length 659;
Best Local Similarity 30.1%; Pred. No. 2.5e-28;
Matches 138; Conservative 67; Mismatches 153; Indels 101; Gaps 18;

QY 8 VKADVAQSSYGLYGQGIQVAVADTGLDTRNDSSMHEAFRGKITALY-ALGRTNANDTN 66
Db 145 IGADTVNLSGLYDGGVVAIVDTGIDAN-----HPDLKGVIGWYDAVNGRSTPYDDQ 198
QY 67 GHGTHVAGSVLNGSTNK---GMAPOANLVFQSIM---DSGGGLGLPSNLQTLFSQAYS 121
Db 199 GHGTHVAGSVLNGSTNKGAPQANLVFQSIMDSGGGLGLPSNLQTLFSQAYS 258
QY 122 GARI-----HTNSMGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPGGT 169
Db 259 GIRVNLGLSSQSSDGTSLSLQAVNNAWDA-----GIVVCAAGNSGPNITYT 306
QY 170 ISAPGTAKNAITVGATENLRPSFGSYADNINHVAQFSRGTGDKRIKPDVMAFGTILS 229
Db 307 VGSFAAASKVITVGA-----VDSNDNIASFSSRGFTADGRLEKPEVVAFGVDIIA 355
QY 230 ARSSLAPDSSFWANHDSKYAYMGGTSMATPIVAG-NVAQLREHFVKNRGITPK-PSLLK 286
Db 356 PRAS---GTSMGTLINDYYTKASGTSNATPHVSGVGAIILOAH-----PSWTPDKVK 404
QY 287 AALIAGA-----ADIGLYPNGNGWGRVTLDKSL---NVAYNSSSLSTSQKATY 335
Db 405 TALIETADIVAPKEIADIAGA-----GRVNVYKAIKYDDYAKLFTGTVADKGSATH 457
QY 336 SFTATAGKPLKISLVMSDAPASTTASVTLVNDLVLITAPNGTOYVGNDFTSFYNDNWDG 395
Db 458 TFDVSGATFTATLYWD-----TGSSDIDLXYDPNGNE-VDYSTAYY----- 500
QY 396 RNNVENVFNAPOSQGTITIEVQAYNVFVGPOTFSLAIVN 434
Db 501 --GFEKVGYYNPTAGTWTVKVVSYK---GAANYQVDVVS 534

RESULT 15
US-09-445-472-12
; Sequence 12, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOOO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151965/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Thermococcus celer
US-09-445-472-12
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:46:23 ; Search time 39 Seconds
(without alignments)
2349.757 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSGYGLY.....EVOAYNPVGPQTSLAIVN 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2247	100.0	434	10	US-09-985-689A-1
2	2247	100.0	434	15	US-10-385-662-2
3	2191	97.5	434	10	US-09-985-689A-2
4	2143	95.4	434	10	US-09-985-689A-6
5	2125.5	94.6	433	10	US-09-985-689A-7
6	1998.5	88.9	433	10	US-09-985-689A-5
7	1984.5	88.8	433	10	US-09-985-689A-3
8	1987.5	88.5	433	10	US-09-985-689A-4
9	1581.5	70.4	345	14	US-10-336-324-10
10	1581.5	70.4	345	14	US-10-403-105-13
11	452.5	20.1	659	13	US-10-090-624-12
12	414	18.4	412	13	US-10-090-624-1
13	414	18.4	522	13	US-10-090-624-4
14	414	18.4	654	13	US-10-090-624-16
15	366	16.3	1079	14	US-10-112-488-39

16	363.5	16.2	1208	14	US-10-156-761-13251	Sequence 13251, A
17	346.5	15.4	1139	14	US-10-156-761-10856	Sequence 10856, A
18	338	15.0	1237	14	US-10-314-657-4	Sequence 4, Appli
19	306.5	13.6	519	15	US-10-084-846A-114	Sequence 114, Appl
20	306.5	13.6	13725	15	US-10-084-846A-4	Sequence 4, Appli
21	304.5	13.6	1398	13	US-10-090-624-6	Sequence 6, Appli
22	283	12.6	580	10	US-09-927-827-55	Sequence 55, Appli
23	280.5	12.5	595	10	US-09-927-827-59	Sequence 59, Appli
24	276	12.3	418	9	US-09-966-921A-2	Sequence 2, Appli
25	270	12.0	1101	14	US-10-156-761-12934	Sequence 12934, A
26	251	11.2	280	14	US-10-309-812-2	Sequence 2, Appli
27	251	11.2	397	10	US-09-779-334A-5	Sequence 5, Appli
28	247	11.0	271	10	US-09-813-408-2	Sequence 2, Appli
29	246.5	11.0	379	10	US-09-813-408-6	Sequence 6, Appli
30	246	10.9	271	14	US-10-242-549-56	Sequence 56, Appli
31	244	10.9	271	14	US-10-242-549-46	Sequence 46, Appli
32	244	10.9	271	14	US-10-242-549-48	Sequence 48, Appli
33	244	10.9	271	14	US-10-242-549-50	Sequence 50, Appli
34	244	10.9	271	14	US-10-242-549-52	Sequence 52, Appli
35	244	10.9	271	14	US-10-242-549-54	Sequence 54, Appli
36	244	10.9	271	14	US-10-242-549-60	Sequence 60, Appli
37	242.5	10.8	627	10	US-09-927-827-60	Sequence 60, Appli
38	242	10.8	271	14	US-10-242-549-44	Sequence 44, Appli
39	242	10.8	271	14	US-10-242-549-58	Sequence 58, Appli
40	240	10.7	269	8	US-08-322-678-10	Sequence 10, Appli
41	240	10.7	269	9	US-09-837-235-16	Sequence 16, Appli
42	240	10.7	269	9	US-09-060-854B-6	Sequence 6, Appli
43	240	10.7	269	9	US-09-975-139-1	Sequence 1, Appli
44	240	10.7	269	9	US-09-976-414-8	Sequence 8, Appli
45	240	10.7	269	10	US-09-736-116-49	Sequence 49, Appli

ALIGNMENTS

RESULT 1

US-09-985-689A-1
; Sequence 1, Application US/09985689A
; Publication No. US2003002351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985,689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-1

Query Match 100.0%; Score 2247; DB 10; Length 434;
Best Local Similarity 100.0%; Pred. No. 6.2e-188;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSGYGLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTN 60

Db 1 NDVARGIVKADVAQSSGYGLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTN 60

QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIIMDSGGGLGLPSNLQTLFSQAYS 120

```

Db      61  |NANDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLPSNLQTLFSQAYS 120
Qy      121 |AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db      121 |AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Qy      181 |TVGATENLRSPSGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240
Db      181 |TVGATENLRSPSGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240
Qy      241 |WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY 300
Db      241 |WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY 300
Qy      301 |PNGNQGWGRVTLDKSLNVAVYVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 360
Db      301 |PNGNQGWGRVTLDKSLNVAVYVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 360
Qy      361 |SVTLVNDLVLITAPNGTOYVGNDFTPSYNDNDGGRNNVENVFINAPOSQGTITIEVQAYN 420
Db      361 |SVTLVNDLVLITAPNGTOYVGNDFTPSYNDNDGGRNNVENVFINAPOSQGTITIEVQAYN 420
Qy      421 |VPVGPQTFSLAIYN 434
Db      421 |VPVGPQTFSLAIYN 434

```

RESULT 2

```

US-10-385-662-2
; Sequence 2, Application US/10385662
; Publication No. US20040002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAEKI, KATSUHISA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938US0
; CURRENT APPLICATION NUMBER: US/10/385,662
; PRIOR FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231
; PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp. KSM-KP43
US-10-385-662-2

```

```

Query Match      100.0%; Score 2247; DB 15; Length 434;
Best Local Similarity 100.0%; Pred. No. 6.2e-188;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
Db      1  NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60

Qy      61  NANDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLPSNLQTLFSQAYS 120
Db      61  NANDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLPSNLQTLFSQAYS 120

```

```

Qy      121 |AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db      121 |AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Qy      181 |TVGATENLRSPSGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240
Db      181 |TVGATENLRSPSGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240
Qy      241 |WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY 300
Db      241 |WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY 300
Qy      301 |PNGNQGWGRVTLDKSLNVAVYVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 360
Db      301 |PNGNQGWGRVTLDKSLNVAVYVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 360
Qy      361 |SVTLVNDLVLITAPNGTOYVGNDFTPSYNDNDGGRNNVENVFINAPOSQGTITIEVQAYN 420
Db      361 |SVTLVNDLVLITAPNGTOYVGNDFTPSYNDNDGGRNNVENVFINAPOSQGTITIEVQAYN 420
Qy      421 |VPVGPQTFSLAIYN 434
Db      421 |VPVGPQTFSLAIYN 434

```

RESULT 3

```

US-09-985-689A-2
; Sequence 2, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION: YUJI
; APPLICANT: HATADA, AKINORI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985,689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-2

```

```

Query Match      97.5%; Score 2191; DB 10; Length 434;
Best Local Similarity 96.5%; Pred. No. 4.9e-183;
Matches 419; Conservative 13; Mismatches 2; Indels 0; Gaps 0;

Qy      1  NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
Db      1  NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60

Qy      61  NANDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLPSNLQTLFSQAYS 120
Db      61  NANDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLPSNLQTLFSQAYS 120

Qy      121 |AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db      121 |AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Qy      181 |TVGATENLRSPSGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240
Db      181 |TVGATENLRSPSGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240

```

Qy 241 WANHDSKYA YMGTSKATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALAGAADIGLY 300
Db 241 WANHDSKYA YMGTSKATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALAGAADVGLGY 300
Qy 301 PNGNQGRVTLTKSLNVA YNNESSLSTSKATYSTAGKPLKISLVWSDAPASTTA 360
Db 301 PNGNQGRVTLTKSLNVA YNNESSLSTSKATYSTAGKPLKISLVWSDAPASTTA 360
Qy 361 SVTLVNDLVLITAPNGTQYVGNDFSPYNDNDGRNNVENVFINAPQSGTYTIEVOAYN 420
Db 361 SVTLVNDLVLITAPNGTQYVGNDFSPYNDNDGRNNVENVFINAPQSGTYTIEVOAYN 420
Qy 421 VPVGPQTFSLAIVN 434
Db 421 VPVGPQTFSLAIVN 434

RESULT 4
US-09-985-689A-6
; Sequence 6, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483USO
; CURRENT APPLICATION NUMBER: US/09/985, 689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-6

Query Match 95.4%; Score 2143; DB 10; Length 434;
Best Local Similarity 93.5%; Pred. No. 7, 7e-179;
Matches 406; Conservative 19; Mismatches 9; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAQSSYGLYGGQGI VAVADTGLDTGRNDSMHEAFRGKITALYALGRTN 60
Db 1 NDVARGIVKADVAQSSYGLYGGQGI VAVADTGLDTGRNDSMHEAFRGKITALYALGRTN 60
Qy 61 NANDTNGHGHVAGSVLGNSTKGMAPAQLVFSQIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NANDPNGHGHVAGSVLGNSTKGMAPAQLVFSQIMDSGGGLGGLPSNLQTLFSQAYS 119
Qy 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEPGNGGTISAPGTAKNAI 180
Db 121 AGARIHTNSGAPVNGAYTTDSRVDDYVRKNDMTILFAAGNEPGNGGTISAPGTAKNAI 180
Qy 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKPDYMAPGTFTLSARSSLAPDSSF 240
Db 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKPDYMAPGTFTLSARSSLAPDSSF 240
Qy 241 WANHDSKYA YMGTSKATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALAGAADIGLY 300
Db 241 WANHDSKYA YMGTSKATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALAGAADIGLY 300
Qy 301 PNGNQGRVTLTKSLNVA YNNESSLSTSKATYSTAGKPLKISLVWSDAPASTTA 360
Db 301 PNGNQGRVTLTKSLNVA YNNESSLSTSKATYSTAGKPLKISLVWSDAPASTTA 360

Db 301 PSNGQGRVTLTKSLNVA FVNETSSLSTNOKATYSTAQSGKPLKISLVWSDAPASTTA 360
Qy 361 SVTLVNDLVLITAPNGTQYVGNDFSPYNDNDGRNNVENVFINAPQSGTYTIEVOAYN 420
Db 361 SVTLVNDLVLITAPNGTQYVGNDFSPYNDNDGRNNVENVFINAPQSGTYTIEVOAYN 420
Qy 421 VPVGPQTFSLAIVN 434
Db 421 VPVGPQTFSLAIVN 434

RESULT 5
US-09-985-689A-7
; Sequence 7, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483USO
; CURRENT APPLICATION NUMBER: US/09/985, 689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-7

Query Match 94.6%; Score 2125.5; DB 10; Length 433;
Best Local Similarity 93.5%; Pred. No. 2, 6e-177;
Matches 406; Conservative 19; Mismatches 8; Indels 1; Gaps 1;

Qy 1 NDVARGIVKADVAQSSYGLYGGQGI VAVADTGLDTGRNDSMHEAFRGKITALYALGRTN 60
Db 1 NDVARGIVKADVAQSSYGLYGGQGI VAVADTGLDTGRNDSMHEAFRGKITALYALGRTN 60
Qy 61 NANDTNGHGHVAGSVLGNSTKGMAPAQLVFSQIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NANDPNGHGHVAGSVLGNSTKGMAPAQLVFSQIMDSGGGLGGLPSNLQTLFSQAYS 119
Qy 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEPGNGGTISAPGTAKNAI 180
Db 120 AGARIHTNSGAPVNGAYTTDSRVDDYVRKNDMTILFAAGNEPGNGGTISAPGTAKNAI 179
Qy 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKPDYMAPGTFTLSARSSLAPDSSF 240
Db 180 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKPDYMAPGTFTLSARSSLAPDSSF 239
Qy 241 WANHDSKYA YMGTSKATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALAGAADIGLY 300
Db 240 WANHDSKYA YMGTSKATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALAGAADVGLGF 299
Qy 301 PNGNQGRVTLTKSLNVA YNNESSLSTSKATYSTAGKPLKISLVWSDAPASTTA 360
Db 300 PNGNQGRVTLTKSLNVA FVNETSSLSTSKATYSTAQSGKPLKISLVWSDAPGSTTA 359
Qy 361 SVTLVNDLVLITAPNGTQYVGNDFSPYNDNDGRNNVENVFINAPQSGTYTIEVOAYN 420
Db 360 SVTLVNDLVLITAPNGTQYVGNDFSPYNDNDGRNNVENVFINAPQSGTYTIEVOAYN 419
Qy 421 VPVGPQTFSLAIVN 434


```
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483USO
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: US/09/985.689A
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-4

Query Match      88.5%; Score 1987.5; DB 10; Length 433;
Best Local Similarity 87.3%; Pred. No. 3e-165;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVQAQSSYGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 1 NDVARGIVKADVQAQNNYGLYGQGOVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIQNDSSGGLGGLPSNLQTLFSQAYS 120
Db 61 NASDPNGHGHVAGSVLNG-ALNKGMAPOANLVFQSIQNDSSGGLGGLPSNLQTLFSQAWN 119
QY 121 AGARIHTNSWGAAVNGAYTTDSRVNDDYVRKNDMTILFAAGNEGPNNGGTTISAPGTAKNAI 180
Db 120 AGARIHTNSWGAPVNGAYTANSRQVDEYVRNNDMTILFAAGNEGPNNGGTTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTIFILSARSSLPDSSF 240
Db 180 TVGATENYRPSFGSIADPNPHIAQFSSRGATRDGRIKPDVTAFGTIFILSARSSLPDSSF 239
QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALTAGAADIGLGY 300
Db 240 WANYNSKYAYMGTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALTAGAADIGLGY 299
QY 301 PNGNQGRVTLKSLNVAYVNESSLSSTQKATYSTATAGKPLKISLVMSDAPASTTA 360
Db 300 PNGNQGRVTLKSLNVAYVNEATATAGKATYSTQAGKPLKISLVMTDAPGSTTA 359
QY 361 SVTLVNDLVLITAPNGTQVYGNDFTSYNDNWDGRNNVNFVINAPOSCTYTIIEQAYN 420
Db 360 SYTLVNDLVLITAPNGQKYVGNDFSYPDNNDGRNNVNFVINAPOSCTYTIIEQAYN 419
QY 421 VPGVGPQTFSLAIVN 434
Db 420 VPSGPGQFSLAIVH 433

RESULT 9
US-10-336-324-10
; Sequence 10, Application US/10336324
; Publication No. US20030176304A1
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; APPLICANT: Andersen, Kim
; TITLE OF INVENTION: Protease Variants and Compositions
; FILE REFERENCE: 5349.204-US
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: US/09/512.251A
; PRIOR FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 10
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-10-336-324-10

Query Match      70.4%; Score 1581.5; DB 14; Length 345;
Best Local Similarity 95.3%; Pred. No. 6.9e-130;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVQAQSSYGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 29 NDVARGIVKADVQAQNNYGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 88
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIQNDSSGGLGGLPSNLQTLFSQAYS 120
Db 89 NANDPNHGHVAGSVLNG-ATNKGMAPOANLVFQSIQNDSSGGLGGLPSNLQTLFSQAYS 147
QY 121 AGARIHTNSWGAAVNGAYTTDSRVNDDYVRKNDMTILFAAGNEGPNNGGTTISAPGTAKNAI 180

; SEQ ID NO 13
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-10-403-105-13

Query Match      70.4%; Score 1581.5; DB 14; Length 345;
Best Local Similarity 95.3%; Pred. No. 6.9e-130;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVQAQSSYGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 29 NDVARGIVKADVQAQNNYGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 88
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIQNDSSGGLGGLPSNLQTLFSQAYS 120
Db 89 NANDPNHGHVAGSVLNG-ATNKGMAPOANLVFQSIQNDSSGGLGGLPSNLQTLFSQAYS 147
QY 121 AGARIHTNSWGAAVNGAYTTDSRVNDDYVRKNDMTILFAAGNEGPNNGGTTISAPGTAKNAI 180
```

Db 148 AGARIHTNSWGPVNGVAYTTDSRNVDVYVRKNDMTILFAAGNEGSGCTISAPGTAKNAI 207
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAPGTFILSARSSILAPDSF 240
Db 208 TVGATENLRPSFGSYADNINHVAFSSRGPTDGRIRKPDVMAPGTFILSARSSILAPDSF 267
QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGADIGLY 300
Db 268 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGADVGLGF 327
QY 301 PNGNQGWGRVTLDKSLNV 318
Db 328 PNGNQGWGRVTLDKSLNV 345

RESULT 11
US-10-090-624-12
; Sequence 12, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Thermococcus celer
US-10-090-624-12

Query Match 20.1%; Score 452.5; DB 13; Length 659;
Best Local Similarity 30.1%; Pred. No. 9.4e-31;
Matches 138; Conservative 67; Mismatches 153; Indels 101; Gaps 18;

QY 8 VKADVAQSSYGLYGQCIIVAVADTGLDTRNDSSMHEAFRGKITALLY-ALGRTNNAIDTN 66
Db 145 IGDVTWNSLGYDGGVWVAIVDTGIDAN-----HPDLKGVICGYDAVNGRSTPYDDQ 198
QY 67 GHGTHVAGSVLNGSTNK---GMAPOANLVFQSIM--DSGGGLGLPSNLQTLFSQAYSA 121
Db 199 GHGTHVAGSVLNGSTNK---GMAPOANLVFQSIM--DSGGGLGLPSNLQTLFSQAYSA 121
QY 122 GARI-----HTNSWGAANVNGAYTTDSRNVDVYVRKNDMTILFAAGNEGSGT 169
Db 259 GIRVINLSLGSQSSDGTDLSQLAVNNAWA-----GIVCVAGNSGPNYIT 306
QY 170 ISAPGTAKNAITVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAPGTFILS 229
Db 307 VGSPPAAASKVITVGA-----VDSNDNIAFSRSGPTADGRKLPKPEVAPGVADIIA 355
QY 230 ARSSLAPDSSFWANHDSKYAYMGTSMTATPIVAG-NVAQLREHFVKNRGITPK--PSLLK 286
Db 356 PRAS---GTSMTGPIINDYITKASGTSMTATPIVAGNVAQLREHFVKNRGITPK--PSLLK 404
QY 287 AALIAGA-----ADIGLYPNGNQGWGRVTLDKSL---NVAYVNESSLSSTSQKATY 335
Db 405 TALLETADIVAPKEIADIVAGA-----GRNVYKAIKYDDYAKLITFTGSVADKGSATH 457
QY 336 SFTATAGKPLKISLVWSDAPASTASVTLVNDLVLITAPNGTQVYVGNDFTSYNDNMDGRN 395
Db 458 TFDVSGATEVTATLYWD-----TGSSDIDLVLDPNGNE-VDYSYTAIY-----G 500

QY 396 RNNVENVFINAPSGTYTIEVQAYNVVPQPTFSLAIVN 434
Db 501 --GFEKVGYNPTAGTWTVKVSYK---GAANYQVDWVS 534

RESULT 12
US-10-090-624-1
; Sequence 1, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-10-090-624-1

Query Match 18.4%; Score 414; DB 13; Length 412;
Best Local Similarity 29.8%; Pred. No. 1.1e-27;
Matches 136; Conservative 59; Mismatches 147; Indels 114; Gaps 19;

QY 18 GLYQGGQIVAVADTGLDTRNDSSMHEAFRGKITALLYALGRTNNAI-----DTNQHGH 71
Db 22 GYDGGITIGITDID-----ASHPDLQGGV-----IGWVDFVNGRSYPYDDHGHGH 70
QY 72 VAGSVLNGSTN---KGMAPQANLVFQSIM--DSGGGLGLPSNLQTLFSQAYSAAGARI 125
Db 71 VASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTIKGVEMAVDNKDKYGIKV 130
QY 126 HTNSWGA-----AVNGAYTTDSRNVDVYVRKNDMTILFAAGNEGSGTISAP 173
Db 131 INLSLSSQSSDGTDLSQLAVNNAWA-----GLVWVAAGNSGPNKYITGSP 178
QY 174 GTAKNAITVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAPGTFILSARSS 233
Db 179 AAASKVITVGA-----VDKVDVTSFSSRGPTADGRKLPKPEVAPGVADIIAARAS 227
QY 234 LAPDSSFWANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPK--PSLLKALIA 291
Db 228 ---GTSMGQPIINDYITTAAPGTSMTATPIVAGNVAQLREHFVKNRGITPK--PSLLKALIA 277
QY 292 GA-----ADIGLYPNGNQGWGRVTLDKSLNVAAYVNESSLSSTSQKATY 338
Db 278 TADIVKPEDEIADIVAGA-----GRNVYKAIKYDDYAKLITFTGSVADKGSATH 328
QY 339 ATAGKPLKISLVWSDAPASTASVTLVNDLVLITAPNGTQVYVGNDFTSYNDNMDGRN 398
Db 329 ISGASFVATLYWDNAN-----SDLDLYLDPNGNQ-VDYSYTAIY-----G 369
QY 399 VENNVFINAPSGTYTIEVQAYNVVPQPTFSLAIVN 434
Db 370 FEKVGYNPTAGTWTVKVSYK---GSANYQVDWVS 402

RESULT 13
US-10-090-624-4
; Sequence 4, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/10/090,624
CURRENT FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: 09/445,472
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patentin version 3.0
SEQ ID NO 4
LENGTH: 522
TYPE: PRT
ORGANISM: Pyrococcus furiosus
FEATURE:
NAME/KEY: misc feature
LOCATION: (428)..(428)
OTHER INFORMATION: Xaa at position 428 is Gly or Val.
US-10-090-624-4

Query Match 18.4%; Score 414; DB 13; Length 522;
Best Local Similarity 29.8%; Pred. No. 1.6e-27;
Matches 136; Conservative 59; Mismatches 147; Indels 114; Gaps 19;
QY 18 GLVGGQIVAVATGDTGRNDSMHEAFRGKITALYALGRTNAN-----DTNGHGTH 71
DB 22 GYDGGITIGIITGID-----ASHPDLOQKV-----IGWDFVNGRSYPYDDHGHGTH 70
QY 72 VAGSVLNGSTN---KGMAPQANLVFQSIM--DSGGGLGGLPSNLQTLFSQAYSAGARI 125
DB 71 VASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVWAVDNKDKYGIKV 130
QY 126 HTNSWGA-----AVNGAYTTDSNRVDYVRKNDMTILFAAGNEGPNGGTISAP 173
DB 131 INLSGSSQSSDGTALSOAVNAAWDA-----GLVVVAAGNSGPNKYTIIGSP 178
QY 174 GTAKNAITVGATENLRPSFGSYADNINHVAQFSRGTGDKRIKPDVMAPTGFIILSARS 233
DB 179 AAASKVITVGA-----VDKYDVITFSRSGGTADGRLKPEVVPAGNWIIAARAS 227
QY 234 LAPDSSFANHDSKYAVMGSTMATPIVAGNVACLREHFVKNGITPK--PSLLKAALIA 291
DB 228 ---GTSMGQPINDIYTAAGTSMATPHVAGIAALLQ-----AHPSTWTPDKVKTALIE 277
QY 292 GA-----ADIGLGYPNQGNGRVTLDKSLNAVYNSSSLSTSOKA-----TYSFT 338
DB 278 TADIVKPEIADIAYGA-----GRVNAYKAIN--YDNYAKLVFTGYVANKGSOHQFV 328
QY 339 ATAGKPKISLWSDAPASTTASVTLVNDLVLITAPNGTQYVGNDFTSYNDNWDGRNN 398
DB 329 ISGASFVATLYWDNAN-----SDLDLYLDENGNO--VDYSYTAY-----G 369
QY 399 VENVFIPAQSGTYTIEVQAYNVFVGPTFSLAIVN 434
DB 370 FEKUGYNYPTDGTWIKVWSYS---GSANYQVDVVS 402

RESULT 14
US-10-090-624-16
Sequence 16, Application US/10090624
Publication No. US20020132335A1
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE

FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/10/090,624
CURRENT FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: 09/445,472
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patentin version 3.0
SEQ ID NO 16
LENGTH: 654
TYPE: PRT
ORGANISM: Pyrococcus furiosus
US-10-090-624-16
Query Match 18.4%; Score 414; DB 13; Length 654;
Best Local Similarity 29.8%; Pred. No. 2.2e-27;
Matches 136; Conservative 59; Mismatches 147; Indels 114; Gaps 19;
QY 18 GLVGGQIVAVATGDTGRNDSMHEAFRGKITALYALGRTNAN-----DTNGHGTH 71
DB 154 GYDGGITIGIITGID-----ASHPDLOQKV-----IGWDFVNGRSYPYDDHGHGTH 202
QY 72 VAGSVLNGSTN---KGMAPQANLVFQSIM--DSGGGLGGLPSNLQTLFSQAYSAGARI 125
DB 203 VASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVWAVDNKDKYGIKV 262
QY 126 HTNSWGA-----AVNGAYTTDSNRVDYVRKNDMTILFAAGNEGPNGGTISAP 173
DB 263 INLSGSSQSSDGTALSOAVNAAWDA-----GLVVVAAGNSGPNKYTIIGSP 310
QY 174 GTAKNAITVGATENLRPSFGSYADNINHVAQFSRGTGDKRIKPDVMAPTGFIILSARS 233
DB 311 AAASKVITVGA-----VDKYDVITFSRSGGTADGRLKPEVVPAGNWIIAARAS 359
QY 234 LAPDSSFANHDSKYAVMGSTMATPIVAGNVACLREHFVKNGITPK--PSLLKAALIA 291
DB 360 ---GTSMGQPINDIYTAAGTSMATPHVAGIAALLQ-----AHPSTWTPDKVKTALIE 409
QY 292 GA-----ADIGLGYPNQGNGRVTLDKSLNAVYNSSSLSTSOKA-----TYSFT 338
DB 410 TADIVKPEIADIAYGA-----GRVNAYKAIN--YDNYAKLVFTGYVANKGSOHQFV 460
QY 339 ATAGKPKISLWSDAPASTTASVTLVNDLVLITAPNGTQYVGNDFTSYNDNWDGRNN 398
DB 461 ISGASFVATLYWDNAN-----SDLDLYLDENGNO--VDYSYTAY-----G 501
QY 399 VENVFIPAQSGTYTIEVQAYNVFVGPTFSLAIVN 434
DB 502 FEKUGYNYPTDGTWIKVWSYS---GSANYQVDVVS 534

RESULT 15
US-10-112-488-39
Sequence 39, Application US/10112488
Publication No. US20030082746A1
GENERAL INFORMATION:
APPLICANT: KIKUCHI, Yoshihimi
APPLICANT: DATE, Masayo
APPLICANT: UMEZAWA, Yukiko
APPLICANT: YOKOYAMA, Keiichi
APPLICANT: MATSUI, Hiroshi
TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSGLUTAMINASE
FILE REFERENCE: 219286USOCUNT
CURRENT APPLICATION NUMBER: US/10/112,488
CURRENT FILING DATE: 2002-04-01
PRIOR APPLICATION NUMBER: PCT/JP00/06780
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: JP2000-280098
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: JP11-280098
PRIOR FILING DATE: 1999-09-30
NUMBER OF SEQ ID NOS: 70


```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
; LENGTH: 1079
; TYPE: PRT
; ORGANISM: Streptomyces albogriseolus
US-10-112-488-39

Query Match      16.3%; Score 366; DB 14; Length 1079;
Best Local Similarity 31.6%; Pred. No. 7.1e-23;
Matches 148; Conservative 55; Mismatches 175; Indels 90; Gaps 20;

QY 3 VAR----GIVKADVAQS-----SYLYGQCIQIVAVADTGLDTCGRNDSMHEAPRG 48
Db 160 VARVWLDGVRKASLDSVQIGCTPKAWEAGYDGKGVKIAVLDTGVD-----ATHPDLKG 213
QY 49 KITALYALGRTNANDTNGHGHVAGSVLGNCS-----TNKGMAPOANLVFQSIWDSGGGL 104
Db 214 QVTASKNFTSAPTTGDDVGHGHVAVASIAAGTGAQSKGYKGVAPGAKILNGKVLDDAG-- 271
QY 105 GGLPSNLQTLFQOAYSAGARIHTNSGAAVNGAYTTDSRNVDYVRK--NDMTILF--AA 160
Db 272 FGDDSGILAGMEWAAQAQADIYVMSLG----GMDTPETDPLEAAVDKLSAEKGLFAIAA 327
QY 161 GNEGPNGGTISAPGTAKNAITVGATENLRPSGYSYADNINHVAQPSRGP-TKDGRIKPD 219
Db 328 GNEGPGQ--SIGSPGSADSALTVGA-----VDDKDKLADPSTGPRLGCGAVKPD 374
QY 220 VMAPGTFFILSRSSLAPDSSFWANHDSKYAVVGGTSMATPIVAGNVAOLREHFVKNRGIT 279
Db 375 LTAPGVDITAAAKGNNDIAKEVGEKPEAGYMTISGTSMATPHVAGAAALLKQHPH----- 429
QY 280 PKPSLLKAALIAGAADIGLG-YPNGNQGWGRVTLDKSLNVAIVNNESSLS----- 328
Db 430 WKYAEIKGALTASTKD---GKYTPPEQSGSRVQVDKAITQTIVIAEPVSLSPGVQWPHAD 486
QY 329 ---TSOKATYSFTATAGKPLKISLYWSD-----APAS--TTASVTLVNDLDELITAP-NG 377
Db 487 DKPVTNKLTYRNLTGHEDVTLKLTSTATGPKGAAPAGPFTLIGASTL-----TVPANG 538
QY 378 TQYVGNDFTSYNDNMDGNNYVFINAPQS-----GTYTIEVQAYNV 421
Db 539 TASVDVTADTRLGGAVDGTYSAYVATGAGQSVRTAAAVEREVESYNV 586
```

Search completed: March 10, 2004, 14:51:57
Job time : 40 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:42:47 ; Search time 21 Seconds
(without alignments)
1987.958 Million cell updates/sec

Title: US-09-985-689A-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAYNVPVGQTSLAIWN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	523.5	23.3	1743	2 T18279	multidrug resist
2	497	22.1	1905	2 T18267	multidrug resist
3	347.5	15.5	444	2 B83891	intracellular alka
4	320.5	14.3	442	2 A69587	intracellular alka
5	310.5	13.8	806	2 A41341	microbial serine p
6	302.5	13.5	1398	2 T28159	pyrolysin (EC 3.4
7	283	12.6	580	2 S11890	serine proteinase
8	281	12.5	419	1 S25835	subtilisin (EC 3.4
9	280	12.5	799	2 G83753	subtilisin-type pr
10	279.5	12.4	1345	2 T29090	surface layer-asso
11	277	12.3	420	1 J23407	subtilisin (EC 3.4
12	267	11.9	715	2 JC4908	alkaline serine pr
13	263	11.7	1331	2 A72647	probable surface l
14	259.5	11.5	573	2 C84120	subtilisin-type pr
15	257.5	11.5	513	1 A35742	aqualysin (EC 3.4
16	253.5	11.3	894	2 F69730	cell wall-associat
17	251	11.2	627	2 D75393	serine proteinase,
18	246.5	11.0	402	1 JU0332	alkaline proteinase
19	246.5	11.0	534	1 J50173	alkaline proteinase
20	243	10.8	519	2 S71451	halolysin R4 (EC 3
21	242.5	10.8	401	2 I39974	serine proteinase
22	240	10.7	380	2 A49778	high-alkaline seri
23	240	10.7	488	2 A11930	proteinase [import
24	235.5	10.5	382	1 SUBSN	subtilisin (EC 3.4
25	235	10.5	378	2 A33973	high-alkaline seri
26	235	10.5	382	2 I39780	subtilisin (EC 3.4
27	234	10.4	910	2 C69456	subtilisin sendai
28	234	10.4	1374	2 D72593	hypothetical prote
29	232.5	10.3	525	2 G84406	halolysin [import

30	231	10.3	321	1 S27501	alkaline proteinase
31	230.5	10.3	379	1 SUBSCL	subtilisin (EC 3.4
32	229.5	10.2	601	2 JC4576	serine proteinase
33	227	10.1	1167	1 A35066	streptococcal Csa
34	226.5	10.1	1118	2 H97298	subtilisin like pr
35	225.5	10.0	381	2 JH0778	subtilisin (EC 3.4
36	225.5	10.0	613	2 S75976	hypothetical prote
37	225.5	10.0	1052	2 T17093	intraluminal subti
38	223.5	9.9	381	1 SUBSS	subtilisin (EC 3.4
39	223.5	9.9	381	1 SUBSI	subtilisin (EC 3.4
40	223.5	9.9	381	2 JQ1487	subtilisin (EC 3.4
41	222.5	9.9	275	2 JC1085	subtilisin (EC 3.4
42	222	9.9	384	2 JC4802	alkaline proteinase
43	221.5	9.9	272	2 A23624	subtilisin (EC 3.4
44	221.5	9.9	530	2 A42605	halolysin (EC 3.4
45	221	9.8	1036	2 JC5568	serine proteinase

ALIGNMENTS

RESULT 1

T18279
multidrug resistance transport protein - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
A:Accession: T18279
R:Shaulsky, G.; Loomis, W.F.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z18855
A:Accession: T18279
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1743 <SHA>
A:Cross-references: EMBL:U60086; NID:g1399914; PID:g1399915; PIDN:AAB03331.1
C:Genetics:
A:Gene: tagC

Query Match 23.3%; Score 523.5; DB 2; Length 1743;
Best Local Similarity 27.9%; Pred. No. 2.3e-25;
Matches 165; Conservative 81; Mismatches 150; Indels 195; Gaps 24;

QY	19	LYGOGIVAVADTGLDTR	---NDS-----SMHEAFRGKITALYALGRTNNANDTNGH	68
DB	314	LRGKGILSLADTGLDGHCFDSDKYPILSNVNLNR-KVYTYITTSDDSKVDGH	372	
QY	69	GTHVAGSVLG-----NGSTNKGMAPQANLVFQSIQSGGLGGL--PSNLTQTFQAY	119	
DB	373	GTHICGSAAGTPEDSSVNISSFGLATDAKIAF---FDLASGSSSLTPPSDLKQLYQPLY	429	
QY	120	SAGARIHTNSGA---AVNGAYTDSNVDDYVRKN-DMTILFAAGNEGNGTIS--A	172	
DB	430	DAGARVHCDSWGVSVEGYTGSVSSDTSIDDFLTHPFIILRAAGN---NEQYLSLLT	486	
QY	173	PGTAKNAITVGATENLR-----PSFGSYADNI-----	199	
DB	487	QSTAKNITVGAHTIENHVTIDGPNYVQSSVDINQELICDFDSRYCNVTTAQCLES	546	
QY	200	-----NHVAQSSRGPTKDGRIKPDVWAPGTFIL	228	
DB	547	NATTGLASCCPTLLRKSVIDAANTQPLLNYENNICSFSKSGPTHGDMKFPALVAPGYIT	606	
QY	229	SARSSLA-----PDSSFWANHDSKYVMGTSMTATPIVAGNVAQLREH-----	272	
DB	607	SARNGANTTDCQGGSL-PNTNALLA-IGTSNATSFMAAAATTLAQVLVDGYVPTGSI	664	
QY	273	VKNRGITPKPSLLKAALIAGA-----ADIGLYPNGN-----QGWGRVT	311	
DB	665	VESNKLQPTGSLKALMINNAQLLNGTFLTSSITPSNQVFNFAQSLVQGWGAI	724	
QY	312	LDKSNVAYVNES-----	338	
DB	725	MSNHLFVVNNNNNNNNKNTSGITKFDGIGGLDLRLVKPNQWKEESLSTQNTSYCYTYK	784	

A:Reference:PMID:10678950
A:Accession:U00432
A:Status:Preliminary; translated from GE/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1..1905 <SHA>
A:Cross-references: EMBL:U20432; NID:g664839; PID:g664840; PID:AAA62212.1
C:Genetics:
G:Gene: tagB

RESULT 4

A69587
intracellular alkaline serine proteinase aprX - *Bacillus subtilis*
C/Species: *Bacillus subtilis*
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C/Accession: A69587
R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chi
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galazzi, A.; Galler
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Hoisappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinios
A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauee
Y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetcel
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon
A/Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero
akeuchi, M.; Tanaka, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosaato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamanoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, J.
A/Authors: Yoshikawa, H.F.; Zumschein, E.; Yoshikawa, H.; Danchin, A.
A/Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A/Reference number: A69580; PMID:98044033; PMID:9384377
A/Accession: A69587
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA

A;Residues: 1-442 <KUN>
A;Cross-references: GB:Z99113; GB:AL009126; NID:G2634090; PIDN:CAB13610.1; PID:el183385;
A;Experimental source: strain 168
C;Genetics:
A;Gene: aprX
C;Superfamily: subtilisin homology
F;146-398/Domain: subtilisin homology <SBT>

Query Match 14.3%; Score 320.5; DB 2; Length 442;
Best Local Similarity 29.7%; Pred. No. 3.5e-13;
Matches 107; Conservative 50; Mismatches 104; Indels 99; Gaps 16;

QY 11 DVNQSSGLYGGQGVAVADTGLDTRNDSSMEAFRGKITALYALGRTNAN-----D 64
DB 136 EVVRNQTLTGKGVAVADTGLDTRNDSSMEAFRGKITALYALGRTNAN-----D 64
QY 65 TNGHGTTHVAGSVLGNSTN-----KGMAPQANLVFQIMDSGGGLGSLPNLQTLFQOAYS 120
DB 184 DNGHGTTHVAGSVLGNSTN-----KGMAPQANLVFQIMDSGGGLGSLPNLQTLFQOAYS 120
QY 121 AGARIHTNSGAAVN-----GAYTDSNRVDDYVRKND-----MTIL 157
DB 237 -----EWCIQYNEPNDEPIDIMSMGLGGLALRYDHEQEDPLVRVAVEEAWAGIVVC 288
QY 158 FAAGNEPNGGTTISAPGTAKVAITVGTATENLRPSFGSYADNINHVAFQSGRGTCKDGRK 217
DB 289 VAAGNSQDSTQTTASPOVSEKIVTVGLDNNTA-----SSDDDTVASFSRGTGYVYKEX 344
QY 218 PDVMAPTFTILSRSSSLAPSSP-----WANHSKYAYMGTSMTATPIVAGNVQALREHF 272
DB 345 PDILAPGWNIIISLRS-----PNSYIDKLQKSRVSGSYFTSGTSMATPICAGIAALILQ-- 399
QY 273 VKRGITPK--PSILKAALTAGADIGLGYPNQCGWGRVTLDSLVN-----AVNNESSIL 327
DB 400 -QNPDLTPDEVKELLK-----NGTDK-----KQEDENIYGAGVNAENSV 439

RESULT 5
A11341
Microbial serine proteinase (EC 3.4.21.-), minor (vpr), precursor - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 20-Jun-2000
C;Accession: A41341; B41341; S39700; D69730
R;Stoma, A.; Rufo Jr., G.A.; Theriault, K.A.; Dwyer, M.; Wilson, S.W.; Pero, J.
J. Bacteriol. 173, 6889-6895, 1991
A;Title: Cloning and characterization of the gene for an additional extracellular serine
A;Reference number: A41341; MUD:92041574; PMID:1938892
A;Accession: A41341
A;Molecule type: DNA
A;Residues: 1-806 <GLA>
A;Cross-references: GB:M76590; NID:G413819; PIDN:AAA22881.1; PID:G413820
A;Accession: B41341
A;Molecule type: protein
A;Residues: 161-195 <SL2>
R;Glaser, P.; Kunst, F.; Arnaut, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionescu,
A.; Rapoport, G.; Danchin, A.
Mol. Microbiol. 10, 371-384, 1993
A;Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region fr
A;Reference number: S39655; MUD:95020537; PMID:7934828
A;Accession: S39700
A;Molecule type: DNA
A;Residues: 1-806 <GLA>
A;Cross-references: EMBL:X73124; NID:G413923; PIDN:CAA51501.1; PID:G580871
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Bruchchi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Ewington, J.; Fabbret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
Tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
Rieser, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpatra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUD:98044033; PMID:9384377
A;Accession: D69730
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-806 <KUN>
A;Cross-references: GB:Z99123; GB:AL009126; NID:G2636240; PIDN:CAB15835.1; PID:G2636344
A;Experimental source: strain 168
C;Comment: The amino terminal sequence of the mature protein and a molecular weight of c
C;Genetics:
A;Gene: vpr
C;Start codon: TTG
C;Superfamily: microbial serine proteinase vpr; subtilisin homology
C;Keywords: hydrolase; serine proteinase
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-160/Domain: propeptide #status predicted <PRO>
F;180-548/Domain: subtilisin homology #status atypical <SBT>

Query Match 13.8%; Score 310.5; DB 2; Length 806;
Best Local Similarity 23.5%; Pred. No. 3.4e-12;
Matches 129; Conservative 58; Mismatches 157; Indels 205; Gaps 18;

QY 18 GLYGQGVAVADTGLDTR-----NDSSMEAFRGKITALYALGRTN 60
DB 177 GYTGGIKVAIDTGEYVNHDPDLKNFGQYGYDFVNDYDPKPTPTG-----D 225
QY 61 NANDTNGHGTTHVAGSVLGNSTNKGMAPQANLVFQIMDSGGGLGSLPNLQTLFQOAYS 120
DB 226 PRGEATDGTHTVAGTVAANG--TIKGVAPDATTLLAVRVLPQG--SGTTENVIAQVERAVQ 282
QY 121 AGARIHTNSGAAVNGAYTTDSNRVDDYVRKNDTILFAAGNEPNGGTTISAPGTAKNAI 180
DB 283 DGADVMNLGLNSLNPDPWATSTAL--DWAMSEGVAVVTSNGNSGPNGTGSPGTSREAI 341
QY 181 TVGATE-----NLKPSFGSY-----ADNI----- 195
DB 342 SVGATQLPLNEAVTFFGYSYSAKYNGYNKEDDVKALNNKVELVEAGIGEAKEKDFEGDLT 401
QY 196 -----ADNI----- 199
DB 402 GKAVVYKRGSIATFVDKADNNAKAGAGMVMVNNLSGSEIANVPGMVSPTIKLSLEDEK 461
QY 200 -----NHVAFSSRGPTKD--GRKPDVMAPTFTILSRSSSLAP 236
DB 462 VSALKAGETKTKTFTKLTUSKALGEQVADSSRGVPMDTWMIKPDISAPGVNIVSPTIHPD 521
QY 237 DSSFWANHSKYAYMGTSMTATPIVAGNVQALREHFVKNRGITPKPSL--LKAALIAGAA 294
DB 522 D-----HPYGYGSKQGTSMASPHIAGAVIKQ-----AKPKWSVEQIKAAIMNTAV 568
QY 295 DI-----GLGYPNQCGWGRVTLDSLVNVAAYNNESSLSTSQKATYSFTATAGPKPLISLV 350
DB 569 TLKDSGGEVYFHNAQAGASARI-----WNAIKADSLVSPGYSY-----GTFLK----- 612
QY 351 WSDAPASTTASVTLVNDLVLITAPNGTQYQYGNDFTPSYNDNWDG--RNNVENYFINAPQ 408
DB 613 -ENGNKTNFTIENQ-----SSIRKSYTLEYSFNGSGISGTSRVRVIPAQ 660
QY 409 SGTYTEVQ 417
DB 661 TGKATAKVK 669

RESULT 6
T28159
Pyrolysin (EC 3.4.-.-) - Pyrococcus furiosus
C;Species: Pyrococcus furiosus
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 24-Oct-2000
C;Accession: T28159

R; Voorhorst, W.G.B.; Eggen, R.I.L.; Geerling, A.C.M.; Platteeuw, C.; Siezen, R.J.; Vos, J. Biol. Chem. 271, 20426-20431, 1996
 A:Title: Isolation and characterization of the hyperthermostable serine protease, pyroly
 A:Reference number: Z20481; MUID:96355370; PMID:8702780
 A:Accession: T28159
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1398 <VOO>
 A:Cross-references: EMBL:U55835; NID:g1556462; PID:g1556463; PIDN:AAB09761.1
 A:Experimental source: DSM3638
 C:Genetics:
 A:Gene: pls
 C:Keywords: hydrolase; serine proteinase

Query Match 13.5%; Score 302.5; DB 2; Length 1398;
 Best Local Similarity 26.6%; Pred. No. 2.3e-11;
 Matches 139; Conservative 57; Mismatches 178; Indels 149; Gaps 22;
 QY 21 GGGQIVAVADTGLDTRGNDSD-----SMHEAFRGKITALYALGRTNAN----- 63
 DB 301 GNGYDIYVDDTDLDFDEVPLGQVNTYDVAVFSYTYGPNLYVLABIDPENGVEYAVFGW 360
 QY 64 DTNGHGHVAGSVLNGSGTN-----KGMAPQAN 91
 DB 361 DCHGHGHVAGTVAGVDSNDAWDLMSYSGEVEFSRLYGWDYTNVTTDTVQGVAPGAQ 420
 QY 92 LVFQIMSDGGGLGSPNLQTLFQAYSAGARIHTNSGAVNGAYT--TDSRN--VDD 147
 DB 421 INAIRVLRS--DGRGSMWDIIEGM--TYAATHGADVISMILGG--NAPYLDGTDPSVAVD 476
 QY 148 YVRKNDMTILFAAGNEFGNGGTISAPGTAKNAITVGTATENRPSFGSYAD----- 197
 DB 477 LFEKGVVVFVIAAGNEGPGINIVSGFVATKALTVGAAA--VPINGVVVSQALGYDPDYG 535
 QY 198 -----NINVAOFSSRGPTKGRIPKDPVAPGTFTILSARSSLAPSSFWANHDSKYA 249
 DB 536 FYFFPAYTNV--RIAPSSRGPRIDGEIKPNVAVPGYGYSSLPFWIGGADF----- 585
 QY 250 YMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGAADI-----GLGYPNG 303
 DB 586 -MSGTSMATPHVSGVALLISG--KPPEGIYNPDIKKVLESQATWLGSDPYTGQKYTEL 643
 QY 304 NQGWGRVTLKSLNVAYNNESSLSTSQKATVFTATGKPKLSLWSDAPASTTA--- 360
 DB 644 DOCHGLVNVTKSWEI-----LKAINGTTLPIVDHWADKSYSDFAEYL 685
 QY 361 SYTLVNDLVLITAPN-----GTQYVGN-----DFTSPYNDW-----DG-----RNNVNVF 403
 DB 686 GVDVIRGLVARNISIPDIVWEHKKYVDTEYRFEIYATPEWPKPFVSGSVILENNTEFVL 745
 QY 404 -----INAPQSGTY-----TIEVQAYNVVPGPQTF 429
 DB 746 RVKYDVEGLEPLGVGRIIIDDPPTTPVIEDELNTIVIEPKFT 788

RESULT 7
 S11890
 serine proteinase (EC 3.4.21.-) precursor, extracellular - Xanthomonas campestris pv. ca
 N:Alternate names: subtilisin-related proteinase
 C:Species: Xanthomonas campestris pv. campestris
 C:Date: 21-Nov-1993 #sequence_revision 07-Feb-1997 #text_change 03-Dec-1999
 C:Accession: S11890
 R:Jiu, Y.N.; Tang, J.L.; Clarke, B.R.; Dow, J.M.; Daniels, M.J.
 Mol. Gen. Genet. 220, 433-440, 1990
 A:Title: A multipurpose broad host range cloning vector and its use to characterise an e
 A:Reference number: S11890; MUID:90251253; PMID:2187155
 A:Accession: S11890
 A:Molecule type: DNA
 A:Residues: 1-580 <LIU>
 A:Cross-references: EMBL:X51635; NID:g48533; PIDN:CAA35962.1; PID:g48534
 A:Experimental source: Xanthomonas campestris pv. campestris
 A:Note: the sequence from Fig. 4 is inconsistent with that from Fig. 3 in having 205-Ala
 C:Superfamily: subtilisin; subtilisin homology

C:Keywords: extracellular protein; hydrolase; serine proteinase
 F:1-32/Domain: signal sequence #status predicted <SIG>
 F:168-423/Domain: subtilisin homology <SBT>

Query Match 12.6%; Score 283; DB 2; Length 580;
 Best Local Similarity 26.2%; Pred. No. 1.3e-10;
 Matches 127; Conservative 55; Mismatches 147; Indels 156; Gaps 24;
 QY 21 GGGQIVAVADTGL-----DTGRNDSSMHEAFRGKITALYALGRTNAN----- 64
 DB 168 GSGTVVAVIDTGITSHADLNANILAGYDFISDATTARDGNGRDSNAADGDMYAAECGA 227
 QY 65 -----TNGHGHVAGSVLNGSGTNKGMAPQ-----NLVFGSIMD 99
 DB 228 GIPAASSWHGHVAGTVAAVNTTGVAGTAGYAKVVPVVLGKGGSLDIADAIWA 287
 QY 100 SGGGLGGLPSNLQ--TLFSQAYSAGARIHTNSGAAVNGAYTTDSRNVDYVVRKNDMTIL 157
 DB 288 SGGTVSGIPANANPAEIVINMSLGGGSCSTTQON--AINGAVSRGT-----TVV 334
 QY 158 FAAGNEGPGTISAPGTAKNAITVGTATEN--LRPSFGSYADNINHVAQFSSRGPTKGR 215
 DB 335 VAAGNDASNVSG--SLPANCANVIAAATTSAGAKASYSNFGTGI----- 377
 QY 216 IKDPVNAPTFTILSARSS--LAPDSSFANHDSKYAYMGGTSMATPIVAGNVAQLREHFV 273
 DB 378 ---DVSAPOSSILSTLNSGTTTPGS-----ASYASYNGTSMASPHVAGVVALVQS--V 425
 QY 274 KNRGITPK--PSLLK--AALIAGAADILGYPNGOGWGRVTLKSLNVAYNNESS----- 325
 DB 426 APTALTFAAVETLLKNTARALPGAC-----SGCGGAGIVNADAATAA--INGSGGGG 477
 QY 326 -----SLSTSQKATVFTATGKPKLSLWSDAPASTTASVTL-----VND 367
 DB 478 GGGVTLNGTPTVTVGLGAATGAELNYITVPAG-----SGTLTVTTSGGSGD 523
 QY 368 LDVLI---TAPNGTQVGNDFTPYNDNWDGRNVENVFINAPQSGTYTIEVQAYNVVPG 424
 DB 524 ADLYVRAGSAPTDSAYT---CRPYRS-----GNAETCTITAP--SGTYVYRLKAYS----- 569
 QY 425 PQTF 429
 DB 570 --TFS 572

RESULT 8
 S25835
 subtilisin (EC 3.4.21.62) precursor - Bacillus sp. (strain TA41)
 C:Species: Bacillus sp.
 C:Date: 22-Nov-1993 #sequence_revision 20-Feb-1995 #text_change 18-Jun-1999
 C:Accession: S25835
 R:Davall, S.; Feiler, G.; Narinx, E.; Gerday, C.
 Gene 119, 143-144, 1992
 A:Title: Sequence of the subtilisin-encoding gene from an antarctic psychrotroph Bacill
 A:Reference number: S25835; MUID:93012966; PMID:1398082
 A:Accession: S25835
 A:Molecule type: DNA
 A:Residues: 1-419 <DAV>
 A:Cross-references: EMBL:X63533; NID:g40198; PIDN:CAA45096.1; PID:g40199
 C:Superfamily: subtilisin; subtilisin homology
 C:Keywords: extracellular protein; hydrolase; serine proteinase
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-110/Domain: propeptide #status predicted <PRO>
 F:111-419/Product: microbial serine proteinase #status predicted <MAT>
 F:135-373/Domain: subtilisin homology <SBT>
 F:144,184,359/Active site: Asp, His, Ser #status predicted

Query Match 12.5%; Score 281; DB 1; Length 419;
 Best Local Similarity 33.0%; Pred. No. 1.1e-10;
 Matches 87; Conservative 38; Mismatches 97; Indels 42; Gaps 13;
 QY 21 GGGQIVAVADTGLDTRGNDSSMH-EAFRGKITALYALGRT---NNANDTNGHGHVAGSV 76

Db 135 GAGINIAVLDTGVTNHPDLNNEVECKD-----FTVTGNTFTNSCTDQGHGTHVAGSA 189
QY 77 LGNGSTNK---GMAPOANL-VFOSIMDSGGGLG-GLPSPNLQTLFQAYAGARIHTN-SW 130
Db 190 LANGGTGSGYGVYAPADLWYKVLGDDSGYADDAEIRHAGDQATALTNTKVINMSL 249
QY 131 GAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNPGNGGTISAPGTAKNAITVGTATENLRP 190
Db 250 GSGGESSLIT---NAVYAYDKGLVLIATAAGNSGPKSGISYPCALVNAVAALALENTIQ 306
QY 191 SFGSYADNINHVAFQSSRGFTKDG-----RIKPDVMAPGTFTLSARSSSLAPDSSFWANH 244
Db 307 N-GTY-----RVADFSSRGHRTAGDVIQKGVDSAPGAAYST-----W-F 348
QY 245 DSKYVNGGTSMTPIVAGNVAQL 268
Db 349 DGGYATISGTSMAASPHAGLAAKI 372
RESULT 9
G83753
subtilisin-type proteinase (EC 3.4.21.-) vpr precursor [similarity] - Bacillus halodurans
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: G83753
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: G83753
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-799 <SFO>
A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BA04550.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: vpr
C:Superfamily: microbial serine proteinase vpr; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:1-29/Domain: signal sequence #status predicted <SIG>
Query Match 12.5%; Score 280; DB 2; Length 799;
Best Local Similarity 24.7%; Pred. No. 3e-10;
Matches 117; Conservative 50; Mismatches 135; Indels 172; Gaps 16;
QY 18 GLXGQGIIVADVADTGLDTRNDSSMEAFRGKITALVALGRTNANDT-----NGH 68
Db 171 GYTGGITVAITLDGVDYTHPD--LVHAF-GDYKGDWFDIDNDDPQETPPGDPGRIETH 227
QY 69 GTHVAGSVLNGSTNKGMAPQANLVFOSIMDSGGGLGSPNLQTLFQAYAGARIHTN 128
Db 228 GTHVAGTVAANGLT-KGVAPDANLLAYRVLPGG--RGSTAGVIAGTERAVQDQADIMNL 284
QY 129 SWGAANGAYTTDSRNVDDYVRKNDMTILFAAGNPGNGGTISAPGTAKNAITVGTATENL 188
Db 285 SLGNTLNDPDPATSIAL-DWMAEGVAVTNSGSGNNWTVGSPGTSRDAISVGAT--- 340
QY 189 RPSFGSY-----ADNINH----- 201
Db 341 RLPYKYNKASVFTSDGIDYPSADIMGPSPDELELDGETVEYAFAGLGRKPGDFGVDE 400
QY 202 -----VAQSSSRGPT-KDGRIKPDVMAPGTFTLSARSSSLAPD 237
Db 401 GKIALIVRGEIPFVEKAENAKAGAVGAILIYNNVAGVQPTVPVGLAIPITMLNSDEGLKMR 460
QY 202 -----VAQSSSRGPT-KDGRIKPDVMAPGTFTLSARSSSLAPD 237
Db 461 NELENGQNTVFTSIEFDKLVGETVADFSRGPVHWTWIKPDVSAFGVAIVSTPIPHQPD 520
QY 238 SSFWANHDSKYAYMGTSMTPIVAGNVAQLREHFVKNRGITPKPSLLKAALTAGAADI- 296
Db 521 DPY-----GYGSRGQTSMSAPHYVAGAAALLLEAH-PNWGV-----DHVKAALMNTAENLV 569

QY 297 ---GLGYPNNGQGRVTLDKSLNAVYVNESSLSSTQKATY-SFTATAGPLK 346
Db 570 DENGNYRHNTQGAG-----SIRVDAISETLVTPGSHSFGFTTKERGRQVE 617
RESULT 10
T29090
surface layer-associated STABLE proteinase - Staphylothermus marinus
N:Alternate names: hyperthermostable proteinase
C:Species: Staphylothermus marinus
C>Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T29090
R:May, J.; Lupas, A.; Kellermann, J.; Eckerskorn, C.; Baumeister, W.; Peters, J.
Curr. Biol. 6, 739-749, 1996
A>Title: A hyperthermostable protease of the subtilisin family bound to the surface layer
A:Reference number: Z20559; MUID:96385442; PMID:8793300
A:Accession: T29090
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1345 <MAY>
A:Cross-references: EMBL:U57968; NID:g1374755; PID:g1374756; PIDN:AA02323.1
A:Experimental source: strain F1
C:Function:
A:Description: probably serves an exodigestive function related to the organism's energy
A>Note: stoichiometric S-layer component
Query Match 12.4%; Score 279.5; DB 2; Length 1345;
Best Local Similarity 29.9%; Pred. No. 6.5e-10; Indels 81; Gaps 14;
Matches 99; Conservative 45; Mismatches 106; Indels 81; Gaps 14;
QY 46 FRGITALVALGRTNANDTNGHGTVA-----GSVL-----GNGSTNK--GMAPQANLV 93
Db 445 VQGRYAL-----VSDFHGHGTSVATVIAASGRVLDYLDGKLYRIMGVAPGAKI- 495
QY 94 PQSIMDSGGGLGSPNLQTLFQAYAG-----DYVRKNDMTILFAAGNPGNGGTISAPGTAK 177
Db 496 -----AGGDWILGNILVL--EAWLAGFNIVTEEDGYVYLSLDPGPHRADIIISNW 546
QY 131 GAAVNGAYTTDSRNV-----DVRKNDMTILFAAGNPGNGGTISAPGTAK 177
Db 547 GSIYINFWLQFPFGIDYRSSFMDEILAIRNVLIGHVTIVFAAGNPGNGGTISAPGTAK 606
QY 178 NAITVGATE--NLRPSPG---SYADNINHVAFQSSRGFTKDGRIKPDVMAPGTILSARS 232
Db 607 LVITAGASTLDYTRIYGYPEGYAD---EVTFSSSRGPTGQGYPKPDIVNIGAFEWASTR 663
QY 233 SLAPDSSFWANHDSKYAYMGTSMTPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAIAG 292
Db 664 TI-DGRGVAQPD-----VFGTSEATPYTSGTLALIVQAYKEVNTTDPVTAKILLKSS 718
QY 293 AADIGLYPNNGQGRVTLDKSLNAVYVNE 323
Db 719 AKDI--WYPAFSGSGRVDKALKAADTVFISE 747
RESULT 11
S23407
subtilisin (EC 3.4.21.62) 1 precursor - Bacillus sp. (strain TA39)
C:Species: Bacillus sp.
C>Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 18-Jun-1999
C:Accession: S23407
R:Narinx, E.; Davail, S.; Feller, G.; Gerday, C.
Biochim. Biophys. Acta 1131, 111-113, 1992
A>Title: Nucleotide and derived amino acid sequence of the subtilisin from the antarctic
A:Reference number: S23407; MUID:92256481; PMID:1581352
A:Accession: S23407
A:Molecule type: DNA
A:Residues: 1-420 <NAR>
A:Cross-references: EMBL:X62369; NID:g40200; PIDN:CAA44227.1; PID:g40201
C:Genetics:
A:Gene: sub
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: extracellular protein; hydrolase; serine proteinase

F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-111/Domain: propeptide #status predicted <PRO>
F;112-420/Product: microbial serine proteinase #status predicted <MAT>
F;136-374/Domain: subtilisin homology <SBT>
F;145,185,360/Active site: Asp, His, Ser #status predicted

Query Match 12.3%; Score 277; DB 1; Length 420;
Best Local Similarity 31.2%; Pred. No. 28-10;
Matches 98; Conservative 44; Mismatches 120; Indels 52; Gaps 16;

QY 21 GCGQIVAVADTGLDTRNDSSMHEAPRGKITAL--VALGRT---NNANDTNGHGHVHAGS 75
DB 136 GCGINFAVLDTGWNIN-----HPDLRNNVECKDFTVGTITNNSTCDRQGHGHVHAGS 189

QY 76 VLNGSTNK---GMAPQANL-VFQSIMDSGGGLG-GLPSNLQTLFSQAYSAGARHTN-S 129
DB 190 ALADGGTNGVGVAPADADLWAYKVLGDDGSGYADIAAIAIRHAGDQATALNTKVINMS 249

QY 130 WGAANVGAYTDSRNVDDYVRKNDMTILPAAGNEGPNGGTISAPGTAKNAITVGATENLR 189
DB 250 LGSSGESSLITNAVN---YSYKGVLLIAAGNSGPGYQSGIGPGALYNVAVALEN-K 305

QY 190 PFGSVYADNHNVAOPSSRGPT-KDG-----RIKPDVWAPGTFILSARSSLAPDSSFVAN 243
DB 306 VENGTY-----RVADFSRGYSWTDGDAYIQKGDVEISAPGAATYST-----W-- 348

QY 244 HDSKYAYMGSTMATPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGADIGLYPNG 303
DB 349 FDGGYATISGTVASPHAGLAAKIWAQYPSASNVVDVRELQYRAY---ENDILSGYVAG 405

QY 304 -----NQWGRVTL 312
DB 406 YGDDFASGRGFATV 419

RESULT 12
JC4908
alkaline serine proteinase (EC 3.4.-.-) I precursor - Alteromonas sp.
N/Alternate names: subtilase
C/Species: Alteromonas sp.
C/Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 08-Oct-1999
C/Accession: JC4908
R/Tsujibori, H.; Miyamoto, K.; Tanaka, K.; Kaidzu, Y.; Imada, C.; Okami, Y.; Inamori, Y.
BioSci. Biotechnol. Biochem. 60, 1284-1288, 1996
A/Title: Cloning and sequence analysis of a protease-encoding gene from the marine bacterium *Alteromonas*
A/Reference number: JC4908; MUID:97141200; PMID:8987544
A/Accession: JC4908
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-715 <TSU>
A/Cross-references: DDBJ:D38600; NID:91536787; PIDN:BAA18912.1; PID:d1019647; PID:g21602
A/Experimental source: strain O-7
C/Comment: This enzyme belongs to class I subtilisin-like family. It is a chelator-sensitized metalloprotease.
C/Genetics:
A/Gene: aprI
C/Superfamily: subtilisin homology
C/Keywords: hydrolase

F;1-40/Domain: signal sequence #status predicted <SIG>
F;41-150/Domain: amino-terminal propeptide #status predicted <ATP>
F;151-496/Product: alkaline serine protease I #status predicted <MAT>
F;182-452/Domain: subtilisin homology <SBT>
F;497-715/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F;239-294,335-372,478-481/disulfide bonds: #status predicted

Query Match 11.9%; Score 267; DB 2; Length 715;
Best Local Similarity 25.9%; Pred. No. 1.8e-09;
Matches 124; Conservative 51; Mismatches 168; Indels 136; Gaps 23;

QY 21 GCGQIVAVADTGLDTRNDSSMHEAPRGKITALVALGRTNAND-----TNG---- 67
DB 182 GCGVAVLDTGVRPHLDLNDANILPGYDMISNTFTVANDGGARDNARDGDAVTGECCT 241

QY 68 -----HGTHVAG---SVLNGSTNKGMAPQANLVFQSIMDSGGGLGLP 108

Db 242 DSSGQPVPRADQSSHHGTHVAGTAAVTNNGEGVAGVADAKVVPVRVL---GKCGGLT 298
QY 109 SNLQTLFSQAYSAGARHTNNSWGAAY-----NGAYTTDSRNVDDYVRKNDMTILFAAG 161
DB 299 SDIADGIIWASGGSDRVPANANPAVINMSLGGGACSAATTQAINQARNNGTVIVIAAG 358
QY 162 NEGPNGGTISAPGTAKNAITVGAT--ENLRPFGSVYADNHNVAOPSSRGPTDKGRIPKPD 219
DB 359 NNDNNSANYN-PGNCNGVNVVASVGRDGRAYYSNYGANI-----D 398

QY 220 VMAPGTFILSARSSLAPDSSFWANHDS-----KYAYMGSTMATPIVAGNVAQLR-- 269
DB 399 VAAPG-----CAQSFADDPPEGILSTHNSGGSPNSDSYHSQGSTMAAPHVAGVAALIKQA 454

QY 270 -----BHFVN--RGITPXPSSLKALIAAGADI--GLG-----YPNGNGWGRVTL 312
DB 455 KPSATPDEVETILKNTTRSFAGSCNCGTGVVDAANAALGDDVVTPTGN-----TL 508

QY 313 DKSINVAVNVNESSLSSTQKATYSFTATAGKPLKISLVMSDAPASTASVTL---VNDLD 369
DB 509 ED--GVAKTGLSAGASNQ--FTFDVPAGK-----TNVFTMSGGTGDAD 550

QY 370 LVITAPNGTYGVNDFTSPYNDNWGR---NNVENVFINAPQSGTYTIEVQAYNPVVG 424
DB 551 LYVKL--GSQ-----PTSSYDCRPYEGGNAEVCSEFDAPQAGTYHVMMINGYKAYS 599

RESULT 13
A72647
Probable surface layer-associated STABLE proteinase APE0607 - Aeropyrum pernix (strain A72647)
C/Species: Aeropyrum pernix
C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C/Accession: A72647
R/Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Taka-awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; DNA Res. 6, 83-101, 1999
A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyrum*
A/Reference number: A72450; MUID:99310339; PMID:10382966
A/Accession: A72647
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1331 <KAW>
A/Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BAA79577.1; PID:d1043363; PID:g51
A/Experimental source: strain K1
C/Genetics:
A/Gene: APE0607

Query Match 11.7%; Score 263; DB 2; Length 1331;
Best Local Similarity 23.6%; Pred. No. 7.3e-09;
Matches 111; Conservative 65; Mismatches 158; Indels 136; Gaps 18;

QY 8 VKADVAOSSYGLYGGQIVAVADTGLDTRNDSSMHE-APRGKITALYA----- 55
DB 346 VYADLS-TAYVLF---LKALSDTGMISGPDPSLLDLSFADETPASPASYGVLEARDFTGD 400

QY 56 -----LGRTNAN-----DTNGH 68

DB 401 GVNDPSAGALAGTYDWVGLLTGESVNLGWRGLDFVAGLVPLGLDQGRWWSILYDTLH 460

QY 69 GTHVAGSVLNGSTN-----KGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQA 118

DB 461 GTSVATVIASRGVFNFLGVITSETSLRGVAPGAKIA-----AGGSF-----LINVFAQL 509

QY 119 YSAG-----ARIHNSWG---AANVGAT--TDSNVDDY-VRKNDMT 155
DB 510 FLSGFEPQDPLNWNVTGEHQVDVNNWSGNSYALRGFLTGADDTATIEDYIVSAGTV 569

QY 156 ILFAAGNEGNGGTISAPGTAKNAITVGATE--NLRPFGSVYADNHNVAOPSSRGPTKD 213
DB 570 IVHAWNGGPGYGTATTPGAGSLIISVGASTLFDYRFFGYLPSGCGDVLISWDRGPSQI 629

QY 214 GRIPDVWAPGTFILSARSSLAPDSSFWANHDSKYAYMGSTMATPIVAGNVAQLREHFV 273

```

Db      630  GVAKPDVNNIGSFANAG-----VPVLGLGNGSLAFIDFGTSEATPWTSGSVALVISAYQ 685
Qy      274  KRGITPKPSLLKAALIAAGADIGLGYPNNGQNGRVRTLDKSLNV-----AY 320
Db      686  QAFGAKPSPLVKAALKSTARD*GA--DAFTQSGQGVVYRAVKAVLEGGVPALSTSVY 743
Qy      321  VNSSSLSTSQKATYSFTATAGKELKISLWSDA--PASTTASVTILVNDL 368
Db      744  ENVYSULLS-----GYSYFFLAPNPVEDTQIYPGVLPGETAVETVLKTL 788

RESULT 14
C84120
subtilisin-type proteinase (EC 3.4.21.-) BH763 precursor [similarity] - Bacillus halodurans
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: C84120
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: C84120
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-757 <STO>
A;Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BA07482.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BR3763
C;Superfamily: microbial serine proteinase vpr; subtilisin homology
C;Keywords: hydrolase; serine proteinase
F;1-20/Domain: signal sequence #status predicted <SIG>

Query Match      11.5%; Score 259.5; DB 2; Length 757;
Best Local Similarity 22.6%; Pred. No. 5.7e-09;
Matches 119; Conservative 58; Mismatches 155; Indels 195; Gaps 20;

Qy      8  VKADVAQSSYGLYGGQIVAVADTGLDTRNDSSMHEAFRG-----KITALLYALG 57
Db     119  VRGMLDEGVHLTKGVKVAVIDTGIDYTHPD--LQSSYKGGYDFVDYDDPMETIASQ 176
Qy     58  RTNNANDTNGHTVAGSVLNGSTKMGAPQANLVFQSIIMDSGGGLGLPSNLQTLFSQ 117
Db     177  -----PFTLHGTHVGIIAANGQV-KGVAPEIYAYRALGPGG--QGTEQVIAAEK 227
Qy    118  AYSAGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAK 177
Db    228  AVEDGVVDVNLSTGNTVNGPDWPTSLALDAAVEGVAVT-SNGNSGPNMTVSGPGTSK 286
Qy    178  NAITVGAT-----EN-----LRPSFGS-----194
Db    287  KAI SVGASAPPLNTPYLTAFGEENEISLYPFGGLPMAFKRDLPMIDVGVGTEKEWGV 346
Qy    195  -----YADINIH-----201
Db    347  AEGKVLIKGMWPFTEKWHVAAXKARGVLIYNNTPGPTGMEGGVNI PVVSIREDG 406
Qy    202  -----VAOFSSRGP--TKDGRIPKPDVMAFGTIFLSARSS 233
Db    407  EFLLEQLQKQNKELTLRTYRKEEDFVALFSRGPVTHTWKPKDVVAFGVS1-----DS 462
Qy    234  LAPDSSFWANHDSKYAVMGTSMTATPIVAGNVAQLREHFVKNGRITPK--PSLLKAALIA 291
Db    463  TIPNNG-----YLGNGTSMAPHVAGAAALIKO-----AHPWTPEQVKALMN 507
Qy    292  GAADI-----GLGYPNNGQNGRVRTLDKSLNVAYVYNESSLSSTSQKATYSFTATAGKPLKI 347
Db    508  TAKKLVDOEGVPHEIQGAGRIQVDRKAV-----AATSLVYPGALSFGK-----551
Qy    348  SLVNSDAPASTASVTL-VNDLDAIVITAPNGTQVGVNDFTSPYNDNW 393
Db    552  ---WSKODLREKRPVVTIENHDTV-----KRTYHISPPFDVPGVGV 591
```

```

RESULT 15
A35742
aqualysin (EC 3.4.21.-) I precursor - Thermus aquaticus
C;Species: Thermus aquaticus
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: A35742; S00620; S00324
R;Terada, I.; Kwon, S.T.; Mayata, Y.; Matsuzawa, H.; Ohta, T.
J. Biol. Chem. 265, 6576-6581, 1990
A;Title: Unique precursor structure of an extracellular protease, aqualysin I, with NH-2
A;Reference number: A35742; MUID:90216674; PMID:2182621
A;Accession: A35742
A;Molecule type: DNA
A;Residues: 1-513 <TER>
A;Cross-references: GB:J90108; GB:D90108; GB:J05414; NID:g217171; PIDN:BA014135.1; PID:g
A;Note: the authors translated the codon CTG for residue 470 as Val, and GGT for residue
R;Kwon, S.T.; Terada, I.; Matsuzawa, H.; Ohta, T.
Eur. J. Biochem. 173, 491-497, 1988
A;Title: Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline serine
A;Reference number: S00620; MUID:88225062; PMID:3286255
A;Accession: S00620
A;Molecule type: DNA
A;Residues: 75-442 <KWO>
A;Cross-references: EMBL:X07734; NID:g48069; PIDN:CAA30559.1; PID:g602091
A;Note: part of this sequence, including the amino and carboxyl ends of the mature prote
R;Matsuzawa, H.; Tokugawa, K.; Hamachi, M.; Mizoguchi, M.; Taguchi, H.; Terada, I.; Kwon
Eur. J. Biochem. 171, 441-447, 1988
A;Title: Purification and characterization of aqualysin I (a thermophilic alkaline serin
A;Reference number: S00324; MUID:88151937; PMID:3162211
A;Accession: S00324
A;Molecule type: protein
A;Residues: 128-170 <MATS>
C;Superfamily: subtilisin; subtilisin homology
C;Keywords: extracellular protein; hydrolase; serine proteinase
F;1-14/Domain: signal sequence #status predicted <SIG>
F;15-127/Domain: propeptide #status predicted <PRO>
F;128-408/Product: aqualysin I #status experimental <MAT>
F;157-364/Domain: subtilisin homology <SPT>
F;255-257,281-283/Region: S1 specificity crevice #status predicted
F;409-513/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F;166,197,349/Active site: Asp, His, Ser #status predicted

Query Match      11.5%; Score 257.5; DB 1; Length 513;
Best Local Similarity 26.0%; Pred. No. 4.5e-09;
Matches 117; Conservative 44; Mismatches 144; Indels 145; Gaps 23;

Qy     16  SYGLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALLY-ALGRTNANDTNGHGTAVAG 74
Db    152  TVTATGRGVNYYVDTGIRT-----THREFGGRARVGYDALG--GNGQDCNGHGTAVAG 203
Qy     75  SVLNGSTNGKMGAPQANLVFQSIIMD--SGGLGLGLPSNLQTLFQAYSAGARIHTN---- 128
Db    204  TI--GGVYTYGVAKAVNLAYRVLDCNGSGTSGTGVAGVDWV-----TRNHRPFAVA 252
Qy    129  --SWGAAVNGAVTTDSRNVDYVRKN---DMTILFAAGNEGPNGGTISAPGTAKNAITVG 183
Db    253  NMSLGGGVSTA-----LDNAVKNSTAGVYVAVAGNDNANACNYS-PARVAEALTVG 304
Qy    184  AT--ENLRPFGSGYADNINNHVAFSSRGPDKGRIPKPDVMAFGTIFLSARSLAPDSFV 241
Db    305  ATTSSDARASFSNYGSCV-----DLFAPGAS1PSA-----W 335
Qy    242  ANHDSKYAVMGTSMTATPIVAGNVAQLREHFVKNGRITP---KPSLLKAALIAAGADIGL 298
Db    336  YTSDDTATQTLNGTSMTATPHVAG-VNAL--YLEQNPSTATPASVAILNGATGRLSGIGS 392
Qy    299  GYPNGNQNGRVRTLDKSLNVAYVYNESSLSSTSQKATYSFTATAGKPLKISLWSDAPAST 358
Db    393  GSPN-----RLLYSLSSGS-----GSTAPCTS 415
Qy    359  ----TASVTLVNDLDAIVITAPNGTQY-----VGNDFTSYNDNWGRN---- 397
```


Db 416 CSYTGSLSGPDYMF---QPNGTYYSFAGTHRAWLRGPACTDF-DLYLWRWDGSRWLT 471

Qy 398 -----NVENFIIINAPQSGTYTIEVQAYN 420

Db 472 VGSSTGPTSEESLSYSCTAGYVLWRIYAYS 501

Search completed: March 10, 2004, 14:46:49
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:39:07 ; Search time 18 Seconds
(without alignments)
1255.469 Million cell updates/sec

Title: US-09-985-689A-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSGYL.....EYQAVNPVGPTFSLAIVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	523.5	23.3	1743	1 TAGC_DICDI	Q23868 dictyosteli
2	497	22.1	1905	1 TAGB_DICDI	P54583 dictyosteli
3	310.5	13.8	806	1 SUBV_BACSU	P29141 bacillus su
4	304.5	13.6	1398	1 PLBV_EYFPU	P72186 pyrococcus
5	283	12.6	580	1 EXPR_XANCP	P23314 xanthomonas
6	277	12.3	420	1 SUBT_BAC99	P28842 bacillus sp
7	257.5	11.5	513	1 AQLI_THEAQ	P08594 thermus aqu
8	253.5	11.3	894	1 WPRB_BACSU	P54423 bacillus su
9	246.5	11.0	402	1 ALP_CEPAC	P29118 cephalospor
10	246.5	11.0	534	1 PROA_VIBAL	P16588 vibrio albi
11	242.5	10.8	401	1 THES_BACSP	Q45870 bacillus sp
12	240	10.7	289	1 SUBS_BACLE	P29600 bacillus le
13	240	10.7	380	1 ELYA_BACAO	P27693 bacillus al
14	240	10.7	380	1 ELYA_BACCS	P41362 bacillus cl
15	239	10.6	269	1 PRMT_BACSP	Q99405 bacillus sp
16	235.5	10.5	382	1 SUBT_BACAM	P00782 bacillus am
17	235	10.5	378	1 ELYA_BACSP	P20724 bacillus sp
18	232	10.3	269	1 SUBE_BACLE	P29599 bacillus le
19	231	10.3	321	1 ISP_EACCS	P29140 bacillus cl
20	230.5	10.3	379	1 SUBT_BACLI	P00780 bacillus li
21	230	10.2	404	1 SMP1_MAGPO	Q9y778 magnaporthe
22	230	10.2	1181	1 SCAL_STRPY	P58099 streptococc
23	227	10.1	1167	1 SCAL_STRPY	P15926 streptococc
24	225.5	10.0	381	1 SUBN_BACNA	P35835 bacillus su
25	225.5	10.0	1052	1 MS1P_CRIGR	Q92248 cricetus
26	225.5	10.0	1052	1 MS1P_HUMAN	Q14703 homo sapien
27	225.5	10.0	1052	1 MS1P_MOUSE	Q9wtz2 mus musculu
28	225.5	10.0	1052	1 MS1P_RAT	Q9wtz3 rattus norv
29	223.5	9.9	381	1 SUBT_BACSA	P00783 bacillus su
30	223.5	9.9	381	1 SUBT_BACST	P29142 bacillus st
31	223.5	9.9	381	1 SUBT_BACSU	P04189 bacillus su
32	221.5	9.9	275	1 SUBT_BACPU	P07518 bacillus pu
33	221.5	9.9	530	1 HLY_HALI7	P29143 halophilic

ALIGNMENTS

RESULT 1

ID	TAGC_DICDI	STANDARD;	PRT;	1743 AA.
AC	Q23868;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Prestalk-specific protein tagC precursor (EC 3.4.21.-).			
GN	TAGC.			
OS	Dictyostelium discoideum (Slime mold).			
OC	Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.			
OX	NCBI_TaxID=44689;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=AX4;			
RX	MEDLINE=97140317; PubMed=8986798;			
RA	Shaulsky G., Escalante R., Loomis W.F.;			
RT	"Developmental signal transduction pathways uncovered by genetic suppressors".			
RL	Proc. Natl. Acad. Sci. U.S.A. 93:15260-15265(1996).			
CC	-!- FUNCTION: Intercellular communication via tagC may mediate integration of cellular differentiation with morphogenesis (By similarity).			
CC	-!- SIMILARITY: In the N-terminal section; belongs to peptidase family S8.			
CC	-!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING			
CC	-!- SIMILARITY: PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.			
CC	-!- SIMILARITY: STRONG, TO TAGS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U60086; AAB03331.1; --			
DR	PIR; T18279; T18279.			
DR	DictyBase; DDB0001795; tagC.			
DR	InterPro; IPR003593; AAA_ATPase.			
DR	InterPro; IPR001140; ABC_TM_transport.			
DR	InterPro; IPR003439; ABC_transporter.			
DR	InterPro; IPR000209; Peptidase_S8.			
DR	Pfam; PF00664; ABC membrane; 1.			
DR	Pfam; PF00005; ABC tran; 1.			
DR	Pfam; PF00082; Peptidase S8; 1.			
DR	PRINTS; PR00723; SUBTILISIN.			
DR	SMART; SM00382; AAA; 1.			
DR	PROSITE; PS50929; ABC_TM1F; 1.			
DR	PROSITE; PS0211; ABC_TRANSPORTER 1; 1.			
DR	PROSITE; PS0893; ABC_TRANSPORTER 2; 1.			
DR	PROSITE; PS50893; SUBTILASE ASP; FALSE_NEG.			
DR	PROSITE; PS00136; SUBTILASE ASP; 1.			
DR	PROSITE; PS00137; SUBTILASE_HIS; 1.			
DR	PROSITE; PS00138; SUBTILASE_SER; FALSE_NEG.			
KW	Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;			

KW Signal. 1 27
 FT CHAIN 28 1743
 FT DOMAIN 316 642
 FT DOMAIN 1450 1687
 FT TRANSMEM 962 982
 FT TRANSMEM 1027 1047
 FT TRANSMEM 1072 1092
 FT TRANSMEM 1157 1177
 FT TRANSMEM 1260 1280
 FT TRANSMEM 1288 1308
 FT ACT_SITE 325 325
 FT ACT_SITE 372 372
 FT ACT_SITE 637 637
 FT NP_BIND 1485 1492
 FT DOMAIN 42 46
 FT DOMAIN 94 103
 FT DOMAIN 643 646
 FT DOMAIN 733 741
 FT DOMAIN 786 792
 FT DOMAIN 1337 1340
 FT DOMAIN 1346 1352
 FT DOMAIN 1353 1357
 FT DOMAIN 1358 1364
 FT DOMAIN 1381 1386
 FT DOMAIN 1707 1729
 FT CARBOHYD 390 390
 FT CARBOHYD 536 536
 FT CARBOHYD 547 547
 FT CARBOHYD 614 614
 FT CARBOHYD 689 689
 FT CARBOHYD 735 735
 FT CARBOHYD 741 741
 FT CARBOHYD 776 776
 FT CARBOHYD 832 832
 FT CARBOHYD 887 887
 FT CARBOHYD 1251 1251
 FT CARBOHYD 1385 1385
 FT CARBOHYD 1386 1386
 FT CARBOHYD 1454 1454
 FT CARBOHYD 1704 1704
 FT SEQUENCE 1743 AA; 194145 MW; 12DB363E2F729839 CRC64;

Query Match 23.3%; Score 523.5; DB 1; Length 1743;
 Best Local Similarity 27.9%; Pred. No. 2.7e-26;
 Matches 165; Conservative 81; Mismatches 150; Indels 195; Gaps 24;

QY 19 LYGGQGVAVADTGLDGR---NDS-----SMHEAFRGKITALYALGRTRNANDTNGH 68
 DB 314 LRKGQILSIATGLDGHCFSDSKYPIPLNSVNLNR-KVTVITTSDDSDKVDGH 372
 QY 69 GTHVAGSVLG-----NGSTNKGMAPQANLVFQSDMGGLGL--PSNLQTLFQOAY 119
 DB 373 GTHICGSAAGTPEDSSVNISFSGLATDAKIAF---FOLASGSSLTFFPSDLKQLYQPLY 429
 QY 120 SAGARHTNSWGA---AVNGAYTTDSRVDDYVRKN-DMTILFAAGNEGNGGTIS--A 172
 DB 430 DAGARVHCDSWGSVSVEGYTGSYSDDTASIDFLFTHPFIILRAAGN---NEQVLSLLT 486
 QY 173 PGTAKNAITVGATENLR-----PSFGSYADNI----- 199
 DB 487 QSTAKNVIIVGAHQTHENYLTGDPNYINYQSSVINQELICDFDSRCYCNVTTAQCCLLES 546
 QY 200 -----NHVAQFSRSGPTKDGRIKPDVMAFGTIL 228
 DB 547 NATTGASCCPTLLRKSVIDAANTOPLLYNENNICFSKSGKTHDGRMKPALVAPGEVIT 606
 QY 229 SARSSILA-----POSSFWANHDSKYAVMGTSMATPTVAGNVAOIAREH-----F 272
 DB 607 SARSNGANITDQCGDGL--PNTNALLA-LSGTSMAISFAAAATTLRLQVLDVGYPTGSI 664
 QY 273 VKNRGITPSPILKALINGA-----ADIGLGYPNGN-----QGWGRVT 311
 DB 273 VKNRGITPSPILKALINGA-----ADIGLGYPNGN-----QGWGRVT 311

Db 665 VESNKLQPTGSLKALMINNAQLNGTFLQITSSITYPSNQVFENFAGASLVQGWGAIR 724
 QY 312 LDKSLNAVAYNESS-----SLSTSQKATYSFT-- 338
 Db 725 MSNWLHVVNNNSNNNNKTSDDGITKFDGIGGLDLRLVXPNQWKESLSGTQNTSYCFTYK 784
 QY 339 -----ATAGKPLK---ISLWSDAPASTTASVTLVNDLILV-----TAPNGT 378
 Db 785 PSSSSNSNNIPRVVATLVWTDPPSYAGAKFNLVNLDLTWIIYRDNGSTIFYSNQGG 844
 QY 379 QYVGNDFTSYNDNWDGRNVNFNAPQSGTYTIEVQAYNVVPGVPTFS 429
 Db 845 SFLG---LAPTQDT---LNNVEGIVHNPTPTMYRWAGTNPVPGPQNF 889

RESULT 2
 TAGB_DICDI STANDARD; PRT; 1905 AA.
 ID TAGB_DICDI STANDARD; PRT; 1905 AA.
 AC P54683; 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Prestalk-specific protein tagB precursor (EC 3.4.21.-).
 GN TAGB.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RC MEDLINE=95262903; PubMed=7744252;
 RA Shaullsky G., Kuspa A., Loomis W.F.;
 RT "A multidrug resistance transporter/serine protease gene is required
 for prestalk specialization in Dictyostelium.";
 RL Genes Dev. 9:1111-1122(1995).
 CC -!- FUNCTION: Inter-cellular communication via tagB may mediate
 integration of cellular differentiation with morphogenesis.
 CC -!- SIMILARITY: In the N-terminal section; belongs to peptidase family
 S8.
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING
 TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.
 CC -!- SIMILARITY: STRONG, TO TAGC.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U20432; AAA62212.1; -.
 DR PIR; T18267; T18267.
 DR MEROPS; S08.UPW; -.
 DR DictyBase; DB0001964; tagB.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR001140; ABC_TM_transpt.
 DR InterPro; IPR003439; ABC_Transporter.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00664; ABC membrane; 1.
 DR Pfam; PF00005; ABC tran; 1.
 DR Pfam; PF00082; Peptidase S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SMC0382; AAA; 1.
 DR PROSITE; PS00929; ABC_TMIF; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 DR PROSITE; PS00136; SUBTILASE ASP; FALSE_NEG.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;
 Signal.

FT SIGNAL 1 31 POTENTIAL.
 FT CHAIN 32 1905 PRESTALK-SPECIFIC PROTEIN TAGS.
 FT DOMAIN 378 700 PROTEASE.
 FT DOMAIN 1518 1756 ABC TRANSPORTER.
 FT TRANSMEM 1011 1031 POTENTIAL.
 FT TRANSMEM 1076 1096 POTENTIAL.
 FT TRANSMEM 1121 1141 POTENTIAL.
 FT TRANSMEM 1210 1230 POTENTIAL.
 FT TRANSMEM 1309 1329 POTENTIAL.
 FT TRANSMEM 1332 1352 POTENTIAL.
 FT ACT_SITE 387 387 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 432 432 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 695 695 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1553 1560 ATP (POTENTIAL).
 FT DOMAIN 63 67 POLY-GLN.
 FT DOMAIN 95 104 POLY-ASN.
 FT DOMAIN 107 134 POLY-ASN.
 FT DOMAIN 311 321 POLY-SER.
 FT DOMAIN 833 837 POLY-SER.
 FT DOMAIN 838 844 POLY-GLY.
 FT DOMAIN 871 876 POLY-LEU.
 FT DOMAIN 1012 1015 POLY-TLE.
 FT DOMAIN 1386 1389 POLY-GLU.
 FT DOMAIN 1398 1404 POLY-GLY.
 FT DOMAIN 1445 1450 POLY-ASN.
 FT DOMAIN 1765 1779 POLY-ASN.
 FT DOMAIN 1782 1785 POLY-SER.
 FT DOMAIN 1807 1812 POLY-PRO.
 FT DOMAIN 1813 1860 POLY-PRO.
 FT DOMAIN 1872 1878 POLY-PRO.
 FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 672 672 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 747 747 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 823 823 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1172 1172 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1522 1522 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1658 1658 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1905 AA; 212518 MW; B8E223FA8B9AE13C CRC64;

Query Match 22.1%; Score 497; DB 1; Length 1905;

Best Local Similarity 28.0%; Pred. No. 1.6e-24;

Matches 162; Conservative 72; Mismatches 162; Indels 182; Gaps 22;

QY 19 LYQGOIVAVADGLDTGR---NDS-----SMHEAFRGKITALVALGRTNANDTNGH 68
 DB 376 LRKGQLSIADTGLDGSCHFFSDSKYIPFNQVNNENHRKVTV---YITHDNEDYVNGH 432
 QY 69 GTHVAGSVLNG-----STKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFQAYSA 121
 DB 433 GTHVCGSAACTPEDSSWAISPSGLATDAKIAFYD-LSSGSGSEPTPEPDSQWXPYLYDA 491
 QY 122 GARIHNSGA-----AVNGAYTDSNRVDYRK-NDMTILFAAGNEGPNGTISAPGTA 176
 DB 492 GARVHDSGMSVSLQGYGYSDDAGIDAFLEYEPFSLRAAGN-NELFASLLAQATA 550
 QY 177 KNAITVGATENLRPSFGS-----YADNI-----199
 DB 551 KNAITVGAETAHNVVSDALEYDFSDNANFORPCLPKKYCNVYATKCCSEVSNVKGL 610
 QY 200 -----NHVAQFSRGTKDGRKPDWAPGTILSARSS- 233
 DB 611 QLCCPASIKQNASDSFTTOPQFNNENNMGSFSGKPTDGRKLPDIPVAPGEVITSARNSG 670
 QY 234 -----LAPDSSFWANDSKYVAGGFSMATPIVAGNVAOLREHF-----VKNRGI 278
 DB 671 ENSTDQGDGL-PPNAGLMSISGISMATPLATAATILRLQVLVDGYPTGESVEENKL 728
 QY 279 TPPELSLKALIAAGADIGLY-----PENGNGWGRVTLDSLVNA 319
 DB 729 LPTGSLIKALMINNAQLNGYFWSASSTNPSNAIPEQINGANLIQGWGLRMN---NWL 785
 QY 320 YNNESS-----SLSTSQKAT-----YSFT-----ATAGK 343

DB 786 YVKSNTPTSPRWIGIGLGNKQKATEWKEDSLSSGLNKSKYCTYKPKSSSSSGGGGT 845
 QY 344 P-LKISLWSDAPASTTASVTLVNDL-----VITAPN--GTQYVGNFTSPYND 391
 DB 846 PRIVATLVWTDPPPSYGAKEFNVLNLLDLSIITIGNSGGSLQAGKVAQP--- 902
 QY 392 NWDGRNVNENFVNAPOSQGYTTEVQAYNVVPGPQTES 429
 DB 903 --DTLNVEGIILNPTKAMVKFTIAGTNVPIGPQKFS 938
 RESULT 3
 ID SUBV_BACSU STANDARD; PRT; 806 AA.
 AC P23141,
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Minor extracellular protease vpr precursor (EC 3.4.21.-).
 GN VPR OR IPA-45R OR BSU38090.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 161-195.
 RX MEDLINE=92041574; PubMed=1938892;
 RA Sloma A., Rufe G.A. Jr., Theriault K.A., Dwyer M., Wilson S.W.,
 RA Pero J.;
 RA "Cloning and characterization of the gene for an additional
 RT extracellular serine protease of *Bacillus subtilis*,"
 RL J. Bacteriol. 173:6889-6895(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95020537; PubMed=7934828;
 RA Glaser P., Kunst F., Arnaud M., Condart M.P., Gonzales W.,
 RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
 RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
 RA Rapoport G., Danchin A.;
 RA "Bacillus subtilis genome project: cloning and sequencing of the 97
 RT kb region from 325 degrees to 333 degrees,"
 RL Mol. Microbiol. 10:371-384(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.F.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
 RA Ghm S.Y., Glaser P., Goffeau A., Golightly E.J., Grand G.,
 RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Heraut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lepidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigic C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rev M., Reynolds S.,
 RA Rieger M., Rivoita C., Roche E., Roche R., Rose M., Sadie F.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitznegger T.,

QY 190 PSFGSYADNINHAOFSSRGPT-KDG-----RIKPDVMPGTFILSARSSLAPDSSFWAN 243
Db 306 VENGTY-----RVADFSSGYSWTGDYAIQGDVEISAFGAAYST-----W-- 348
QY 244 HDSKYAYMGTSMTATPIVAGNVAQREHFVKNRGITPKPSLLKAAALIAGAADIGYGPNG 303
Db 349 FDGGYATISGTSWASPHAAAGLAQWQYPSASNDVVRGELQYRAY---ENDILSGYYAG 405
QY 304 -----NQGWGVTLL 312
Db 406 YGDFASFGGFATV 419
RESULT 7
ID AQL1 THEAQ STANDARD; PRT; 513 AA.
AC P08594;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aqualysin I precursor (EC 3.4.21.-).
GN PSTI.
OS Thermus aquaticus.
OC Bacteria; Deinococcus-Thermus; Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=271;
RN [1] SEQUENCE FROM N.A., AND SEQUENCE OF 15-23.
RC STRAIN=VT1;
RX MEDLINE=90216674; PubMed=2182621;
RA Terada I., Kwon S.-T., Miyata Y., Matsuzawa H., Ohta T.;
RT "Unique precursor structure of an extracellular protease, aqualysin
I, with NH2- and COOH-terminal pro-sequences and its processing in
RT Escherichia coli.";
RL J. Biol. Chem. 265:6576-6581(1990).
RN [2]
RP SEQUENCE OF 75-442 FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=VT1;
RX MEDLINE=88225062; PubMed=3286255;
RA Kwon S.-T., Terada I., Matsuzawa H., Ohta T.;
RT "Nucleotide sequence of the gene for aqualysin I (a thermophilic
RT alkaline serine protease) of Thermus aquaticus YT-1 and
RT characteristics of the deduced primary structure of the enzyme.";
RL Eur. J. Biochem. 173:491-497(1988).
RN [3]
RP SEQUENCE OF 128-170.
RX MEDLINE=88151937; PubMed=3162211;
RA Matsuzawa H., Tokugawa K., Hamakaki M., Mizoguchi M., Taguchi H.,
RA Terada I., Kwon S.-T., Ohta T.;
RT "Purification and characterization of aqualysin I (a thermophilic
RT alkaline serine protease) produced by Thermus aquaticus YT-1.";
RL Eur. J. Biochem. 171:441-447(1988).
CC -!- FUNCTION: Aqualysin I is a thermophilic alkaline serine protease.
CC The optimal temperature for its caseinolytic activity is 80
CC degrees Celsius.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DEVELOPMENTAL STAGE: Secreted from the early stationary phase
CC until the time the cells cease to grow.
CC -!- PTM: THE N- AND C-TERMINAL PRO-SEQUENCES ARE REMOVED THROUGH THE
CC PROTEOLYTIC ACTIVITY OF AQUALYSIN I ITSELF, IN THAT ORDER. THE
CC C-TERMINAL PRO-SEQUENCE IS REQUIRED FOR TRANSLOCATION OF THE
CC PROTEASES ACROSS THE OUTER MEMBRANE.
CC -!- PTM: Two disulfide bonds are present.
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; D90108; BAA14135.1; -;
DR EMBL; X07734; CAA30559.1; -;
DR PIR; A35742; A35742.
DR HSSP; P06873; 2PRK.
DR MEROPS; S08.051; -.
DR InterPro; IPR000209; Peptidase S8.
DR InterPro; IPR009020; Protease Inhib.
DR Pfam; PF00082; Peptidase S8; I.
DR PRINTS; PRO0723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 14
FT PROPEP 15 127
FT CHAIN 128 408 AQUALYSIN I.
FT PROPEP 409 513
FT ACT SITE 166 166 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 197 197 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 349 349 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 513 AA; 53913 MW; DDFDFE6D4A50B785 CRC64;
Query Match 11.5%; Score 257.5; DB 1; Length 513;
Best Local Similarity 26.0%; Pred. No. 1.2e-09;
Matches 117; Conservative 44; Mismatches 144; Indels 145; Gaps 23;
QY 16 SYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALY-ALGRTNNANDTNGHGHVAG 74
Db 152 TYTATGGRVNVYIDIGIRT-----THREFGGRVGVYDALG--GNGQDCNGHGHVAG 203
QY 75 SVLNGSTNGMAPQANLVFQSIQMD--SGGLGGLPSNLQTLFQSYAGARLHTN---- 128
Db 204 TI---GGVTYGVAKAVNLYAVRVLDCNGSGSTSGVIAGVDWV-----TRNHRPAVA 252
QY 129 --SWGAAVNGAYTTDSRNVDDYVRKN--DMTILFAAGNEGNGGTTISAPGTAKNAITVG 183
Db 253 NMSLGGGVST-----LDNAVKNISAGVYVAVAGNDNANACNYS-PARVAEALTVG 304
QY 184 AT--ENLRPSFGSYADNINHAOFSSRGPTKGRIKPDVMPGTFILSARSSLAPDSSFW 241
Db 305 ATTSDDARASFSNYGSCV-----DLFAPGASIPSA-----W 335
QY 242 ANHDSKYAYMGTSMTATPIVAGNVAQREHFVKNRGITP---KPSLLKAAALIAGAADIGL 298
Db 336 YTSDDATQTLNGTSMATPHVAG--VAAL--YLEONPSATPASVASAILNGATTGRLSGIGS 392
QY 299 GYPNGNGGGRVTLDKSLNVAVYVNESSLSSTSQATYSFTATAGKPLKISLVMSDAPAST 358
Db 393 GSPN-----RLLYSLSSGS-----GSTAPCTS 415
QY 359 ----TASVTLVNDLVLITAPNGTQY-----VGNDFTSPYNDNWDGRN--- 397
Db 416 CSYTYTSLSGFGDYNF---QPNGTYYTSPAGTHRAWLRGPAGTDF-DLXLWRWDGSRWLT 471
QY 398 -----NVENVFINAPQSGTYTIEVOAYN 420
Db 472 VGSSTGFTSESLSYSGTAGYLLWRIYAYS 501
RESULT 8
WPRA BACSU
ID WPRA BACSU STANDARD; PRT; 894 AA.
AC P54423; O06726;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Cell wall-associated protease precursor (SC 3.4.21.-) [Contains: Cell
DE wall-associated polypeptides CWBP23 and CWBP52].
GN WPRA OR BSU10770.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]

SEQUENCE FROM N.A., AND SEQUENCE OF 32-54 AND 414-428.

STRAIN=169;
MEDLINE=97158234; PubMed=9004506;
Margot P., Karamata D.;
"The wprA gene of *Bacillus subtilis* 168, expressed during exponential growth, encodes a cell-wall-associated protease.";
Microbiology 142:3437-3444(1996).
[2]
SEQUENCE FROM N.A.
STRAIN=168;
MEDLINE=98015415; PubMed=9353931;
Medina N., Vannier F., Roche B., Autret S., Levine A., Sertor S.J.;
"Sequencing of regions downstream of addA (98 degrees) and citG (289 degrees) in *Bacillus subtilis*,";
Microbiology 143:3305-3308(1997).
[3]
SEQUENCE FROM N.A.
STRAIN=168;
MEDLINE=98044033; PubMed=9384377;
Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
Brouillet S., Bruchsi C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
Joris B., Karamata D., Kasahara Y., Klaer-Bianchini M., Klein C.,
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
Presecan B., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Roche B., Roche B., Rose M., Sadaie Y.,
Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
Sekiguchi J., Sekowska A., Sertor S.J., Serror P., Shin B.S., Soldo B.,
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
Takeuchi M., Tanakoshi A., Tanaka T., Terpetra P., Tognoni A.,
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
Viari A., Wambutt R., Wedler E., Wedler H., Weitzengger T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
"The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*,";
Nature 390:249-256(1997).
-!- FUNCTION: NOT YET KNOWN; COULD BE INVOLVED IN PROTEOGLYCAN
DEGRADATION, BY CLEAVAGE OF ITS PEPTIDE BRIDGES.
-!- SUBCELLULAR LOCATION: Cell-wall bound.
-!- PTM: PROCESSED INTO CWBP23 AND CWBP52.
-!- SIMILARITY: Belongs to peptidase family 88.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Y0981; AAC25926.1; -
EMBL; Y09476; CAAT0641.1; -
EMBL; Z99109; CAB12917.1; -
PIR; F69730; F69730.
HSSP; Q45670; LDBI.
MEROPS; S08.004; -
Subtilisin; BGI1846; wprA.
InterPro; IPR000209; Peptidase_S8.
Pfam; PF00082; Peptidase_S8; 1.

RT "Cloning and nucleotide sequences of the complementary and genomic
 RT DNAs for the alkaline protease from *Acremonium chrysogenum*.";
 RL Agric. Biol. Chem. 55:471-477(1991).
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; D00923; BAA0765.1; -
 DR PIR; J00332; J00332.
 DR HSSP; P06873; 2PRK.
 DR MEROPS; S08.UPA; -
 DR InterPro; IPR000209; Peptidase_S8.
 DR InterPro; IPR009020; Protease_inhib.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolyase; Serine protease; Zymogen; Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT PROPEP 21 120 POTENTIAL.
 FT CHAIN 121 402 ALKALINE PROTEINASE.
 FT ACT_SITE 160 160 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 191 191 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 347 347 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SQ SEQUENCE 402 AA; 42099 MW; 8D030CCD42D918E1 CRC64;
 Query Match 11.0%; Score 246.5; DB 1; Length 402;
 Best Local Similarity 30.3%; Pred. No. 4.4e-09;
 Matches 91; Conservative 32; Mismatches 102; Indels 75; Gaps 14;
 QY 21 GGGQIVAVADTGLDTRGNDSSMEAFRGK-ITALYALGRITNNANDTNGHGHVAGSVLGN 79
 DB 151 GSGTYAYVVDGTI-----LESHNEFSGRAITGVNAVGGSN--ADTNGHGHVAGTI--- 199
 QY 80 GSTNKGMAPQANLVFQIMDSGGG-----LGGPLSNLQTLFQAYSAGARI-----HT 127
 DB 200 GGRTYGVAKNTLAVKVFRRGSSSTSIILDGFWNAVNDIINRGQNKAAISMSLOGGV 259
 QY 128 NSWGAANGAYTTDSRVVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAITVG--AT 185
 DB 260 SAFNNAVNTAY--SRGVLSV-----RAGNDQNAANY-S-PASANAALITVGSIAS 306
 QY 186 ENLRPFGSVADNINHVAFSSRGPTKDGRIKPDVMAPFTILSARSLAPDSFWANHD 245
 DB 307 NWARSSFSNYGSVL-----DIFAPGTSILSA-----WIGGN 337
 QY 246 SKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADI---GLGYPN 302
 DB 338 SATNTISGTSMATPHVTGVVLYLQ-----ALEGLITSGAARIALALATTGRVSPNGSGSPN 393
 RESULT 10
 ID PROA_VIBAL STANDARD; PRT; 534 AA.
 AC P16588;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Alkaline serine exoprotease A precursor (EC 3.4.21.-).
 GN PROA.
 OS Vibrio alginolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 CX NCBI_TaxID=663;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=89326126; PubMed=2546861;
 RA Deane S.M., Robb F.T., Robb S.M., Woods D.R.;
 RT "Nucleotide sequence of the Vibrio alginolyticus calcium-dependent,
 RT detergent-resistant alkaline serine exoprotease A.";
 RL Gene 76:281-288(1989).
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M25499; AAA27550.1; -
 DR PIR; J00173; J00173.
 DR HSSP; Q99405; 1MPT.
 DR MEROPS; S08.050; -
 DR InterPro; IPR000209; Peptidase_S8.
 DR InterPro; IPR007280; PPC.
 DR InterPro; IPR009020; Protease_inhib.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF04151; PPC; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolyase; Serine protease; Zymogen; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT PROPEP 22 141 POTENTIAL.
 FT CHAIN 142 534 ALKALINE SERINE EXOPROTEASE A.
 FT ACT_SITE 180 180 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 363 363 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SQ SEQUENCE 534 AA; 55930 MW; 84E96D9C649D4226 CRC64;
 Query Match 11.0%; Score 246.5; DB 1; Length 534;
 Best Local Similarity 23.9%; Pred. No. 6.3e-09;
 Matches 114; Conservative 70; Mismatches 140; Indels 153; Gaps 24;
 QY 7 IVKADVAQSS--YGL-----YGGQIVAVADTGLDTRGNDSSMEAF 46
 DB 137 IVSADANTNAIWLDRIDQRLPLDNNYSANFDTGTAYVDTGV-----NNAHVEF 190
 QY 47 RGKITALLY-ALGRITNNANDTNGHGHVAGSVLGNSTNKGMAPQANLVFQSIM--DSGGG 103
 DB 191 GRSVSGVDFVDNDADASDCNGHGHVAGTI---GGSILYGVAKXNVNLVGVRLVSCSGGS 247
 QY 104 LGGPLSNLQTLFQAYSAGARIHNTNSWGAANGAYTTDSRVVDYVR---KNDMTILFAA 160
 DB 248 TSGVTIAGVDWAANA--SGPSVANMSLGGQSV-----LDSAVQSAVQSGVSFMLAA 298
 QY 161 GNEGPNGGTISAPGTAKNAITVGAT--ENLRPFGSVADNINHVAFSSRGPTKDGRIK 218
 DB 299 GNSNADACNYS-PARVATGVTVGSTTSTDARSSFNWGSV----- 338
 QY 219 DVMAFGTIFILSARSLAPDSFWANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGI 278
 DB 339 DVFAFGSQIKSA-----W--YDGGYKTIISGTSMATPHVAG-VAAL--YLQENSSV 383
 QY 279 TPKPSLLKAALIAAGADIGLVPGNQGWRVTLDKSLNAVYVNESSSLTSQKATYSFT 338
 DB 384 S--PSQVEALIVRAST-----GKVT-----DTRGVNKLILSLT 416
 QY 339 -----ATAGKPL-----KISLVMSDAPASTASVTLLV---ND 367
 DB 417 DADCGQDCGGPDPTDPDPBEGKLTSGVPSGLSGSQVAYVYVDVAGQRLTVQWYSGSD 476
 QY 368 LDLVITAPNGTQVGNDETSPYNDWDQR-----NNVENVFINAQSGCTYTIETVQAYN 420
 DB 477 ADLYLRFG-----AKPTLNAWDCRPFKYGNNETCTVSAQSQSGRIHYMIQSYS 523

```

RESULT 11
THIS_BACSP STANDARD; PRT; 401 AA.
AC Q45670;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE thermophilic serine proteinase precursor (EC 3.4.21.-) (Ak.1
  protease).
OS Bacillus sp. (strain AK1).
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN [1]_TaxID=1409;
RP SEQUENCE FROM N.A.
RX MEDLINE=95085262; PubMed=7993087;
RA Maciver B., McHale R.H., Saul D.J., Bergquist P.L.;
RT "Cloning and sequencing of a serine proteinase gene from a
  thermophilic Bacillus species and its expression in Escherichia
  coli.";
RL Appl. Environ. Microbiol. 60:3981-3988(1994).
RN [2]_
RX X-RAY CRYSTALLOGRAPHY (1.80 ANGSTROMS) OF 122-401.
RP MEDLINE=20057863; PubMed=10588904;
RA Smith C.A., Toogood H.S., Baker H.M., Daniel R.N., Baker E.N.;
RT "Calcium-mediated thermostability in the subtilisin superfamily: the
  crystal structure of Bacillus Ak.1 protease at 1.8-A resolution.";
RL J. Mol. Biol. 294:1027-1040(1999).
CC -|- COPACTOR: Binds 3 calcium ions and 1 sodium ion per subunit.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- MISCELLANEOUS: Has a pH optimum of 8.5, a temperature optimum of
  75 degrees Celsius.
CC -|- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to licenses@isb-sib.ch).
CC -----
CC EMBL: L29506; AAA63688.1; -.
CC PIR: I39974; I39974.
CC PDB: IDBI; 18-NOV-99.
CC MEROPS: S08.009; -.
CC InterPro: IPR000209; Peptidase_S8.
CC InterPro: IPR009020; Protease_inhib.
CC Pfam: PF00082; Peptidase_S8; 1.
CC PRINTS: PR00723; SUBTILISIN.
CC PROSITE: PS00136; SUBTILISIN_ASP; 1.
CC PROSITE: PS00137; SUBTILISIN_SER; 1.
CC Hydrolyase; Serine protease; Zymogen; Metal-binding; Calcium-binding;
  KW Signal; 3D-structure.
  KW Signal; 1 24 POTENTIAL.
  FT PROPEP 25 121
  FT CHAIN 122 401 THERMOPHILIC SERINE PROTEINASE.
  FT ACT_SITE 160 160 CHARGE RELAY SYSTEM.
  FT ACT_SITE 193 193 CHARGE RELAY SYSTEM.
  FT ACT_SITE 347 347 CHARGE RELAY SYSTEM.
  FT METAL 126 126 CALCIUM 1.
  FT METAL 168 168 CALCIUM 3 (VIA CARBONYL OXYGEN).
  FT METAL 169 169 CALCIUM 1.
  FT METAL 171 171 CALCIUM 2.
  FT METAL 179 179 CALCIUM 2.
  FT METAL 184 184 CALCIUM 2.
  FT METAL 186 186 CALCIUM 2 (VIA CARBONYL OXYGEN).
  FT METAL 204 204 CALCIUM 1.
  FT METAL 204 204 CALCIUM 3.
  FT METAL 207 207 CALCIUM 1.
  FT METAL 209 209 CALCIUM 1 (VIA CARBONYL OXYGEN).
  FT METAL 209 209 CALCIUM 1 (VIA CARBONYL OXYGEN).
  FT METAL 211 211 CALCIUM 1 (VIA CARBONYL OXYGEN).
  FT METAL 211 211 CALCIUM 1 (VIA CARBONYL OXYGEN).

FT METAL 297 297
FT METAL 300 300
FT METAL 323 323
FT METAL 258 260
FT DISULFID 127 128
FT TURN 129 132
FT HELIX 135 136
FT TURN 137 139
FT HELIX 140 140
FT TURN 141 144
FT HELIX 145 147
FT TURN 152 153
FT STRAND 155 160
FT TURN 165 166
FT TURN 168 173
FT TURN 174 179
FT TURN 180 183
FT STRAND 184 184
FT HELIX 193 202
FT STRAND 219 224
FT TURN 228 229
FT TURN 234 246
FT TURN 247 248
FT STRAND 251 254
FT HELIX 263 274
FT TURN 275 276
FT STRAND 278 282
FT STRAND 285 285
FT TURN 298 299
FT STRAND 301 306
FT TURN 308 309
FT STRAND 312 312
FT TURN 314 315
FT STRAND 316 316
FT TURN 320 321
FT STRAND 324 327
FT STRAND 331 335
FT TURN 336 338
FT STRAND 339 343
FT HELIX 346 362
FT TURN 363 364
FT HELIX 367 376
FT TURN 377 377
FT STRAND 379 379
FT TURN 382 383
FT STRAND 384 384
FT TURN 385 387
FT STRAND 388 388
FT STRAND 391 392
FT HELIX 395 399
FT TURN 400 401
SQ SEQUENCE 401 AA; 42835 MW; 1C736EFA4A9F256F CRC64;

Query Match 10.8%; Score 242.5; DB 1; Length 401;
Best Local Similarity 30.9%; Pred. No. 8e-09;
Matches 81; Conservative 30; Mismatches 100; Indels 51; Gaps 10;

QY 11 DVAGSSYGLYGGQIVAVADTGLDTGNDSSMHEAPRGKITALLY-ALGRTNANDTNGHC 69
DB 146 DVTKGS-----SQGEIAVIDTGV-----YTHPDLDGKVIKGYDFVNDYDPMDLNHHG 194
QY 70 THVAG---SVLGNSTNGKMAPQANLVFQSIMDSGGGLGCLPSNLQTLFSQAYSAGARIH 126
DB 195 THVAGIAAEETNNATGIAGVAPNTRILLAVRALDRNG--SGLTSDIADIAYADSGAEVI 252
QY 127 TNSWGAANGAYTTDSRVDYVRKNDMTILPAAGNEGNGGTISAPGAKNAITVGATE 186
DB 253 NLSLGC---DCHTTTLENAYVYAWNKGSVVVAAAGNNG--SSTTFPPASVENYIAGVA-- 305
QY 187 NLRPSFGSYADNINHVACFSRGPTRKGRKPDVWAPGTFILSARSLAPDSSFWANHDS 246
DB 306 -----VDQYDLASFSNYGTW-----VDVWAPGVDIVSTITG-----N 338

```


Db 166 ANAMAVGATDQNNRASFQYAGL-----DIVAPGVNVQSTYFG- 205

QY 235 APDSFWANHSKYAMGTSVATPIVAGNVA-----QLREHFVKN 275

Db 206 -----STYASLNGTSVATPHVAGAAALVKQKPNWSNVQIRNH-LKN 246

RESULT 13

ELYA_BACAO

ID ELYA_BACAO STANDARD; PRT; 380 AA.

AC P27693;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Alkaline protease precursor (EC 3.4.21.-).

OS Bacillus alcalophilus.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=1445;

RN [1]

RP STRAIN=PB92;

RC MEDLINE=91282483; PubMed=2059048;

EX van der Laan J.C., Gerritse G., Mulleners L.J.M., van der Hoek R.A.,

RA Quax W.J.;

RT "Cloning, characterization, and multiple chromosomal integration of a

RT Bacillus alkaline protease gene.";

RL Appl. Environ. Microbiol. 57:901-909(1991).

RN [2]

RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS).

RC STRAIN=PB92;

RX MEDLINE=92390330; PubMed=1518788;

RA van der Laan J.C., Teplyakov A.V., Kelders H., Kaik K.H., Misset O.,

RA Mulleners L.J.M., Dijkstra B.W.;

RT "Crystal structure of the high-alkaline serine protease PB92 from

RT Bacillus alcalophilus.";

RL Protein Eng. 5:405-411(1992).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).

RX MEDLINE=93078250; PubMed=1447775;

RA Sobek H., Hecht H.-J., Ahle W., Schomburg D.;

RT "X-ray structure determination and comparison of two crystal forms of

RT a variant (Asn115Arg) of the alkaline protease from Bacillus

RT alcalophilus refined at 1.85-A resolution.";

RL J. Mol. Biol. 228:108-117(1992).

RN [4]

RP STRUCTURE BY NMR OF 112-380.

RC STRAIN=PB92;

RA Martin J.R., Mulder F.A., Karimi-Nejad Y., van der Zwan J.,

RA Mariani M., Schipper D., Boelens R.;

RT "The solution structure of serine protease PB92 from Bacillus

RT alcalophilus presents a rigid fold with a flexible substrate-binding

RT site.";

RL Structure 5:521-532(1997).

CC -!- COFACTOR: Binds 2 calcium ions per subunit.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- SIMILARITY: Belongs to peptidase family S8.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; M65086; AAA22212.1; -.

DR EMBL; A13738; CAA01128.1; -.

DR PIR; A49778; A49778.

DR PDB; 1AH2; 15-APR-98.

DR MEROPS; S08.038; -.

DR InterPro; IPR000209; Peptidase_S8.

DR InterPro; IPR000209; Protease_inhib.

DR Pfam; PF00082; Peptidase_S8; 1.

DR PRINTS; PR00723; SUBTILISIN.

DR PROSITE; PS00136; SUBTILASE ASP; 1.

DR PROSITE; PS00137; SUBTILASE HIS; 1.

DR PROSITE; PS00138; SUBTILASE SER; 1.

KW Hydrolase; Serine protease; Zymogen; Metal-binding; Calcium-binding;

KW Signal; 3D-structure.

FT SIGNAL 1 27 POTENTIAL.

FT PROPEP 28 112

FT CHAIN 113 380 ALKALINE PROTEASE.

FT ACT_SITE 143 143 CHARGE RELAY SYSTEM.

FT ACT_SITE 173 173 CHARGE RELAY SYSTEM.

FT ACT_SITE 326 326 CHARGE RELAY SYSTEM.

FT METAL 113 113 CALCIUM 1.

FT METAL 151 151 CALCIUM 1.

FT METAL 184 184 CALCIUM 1 (VIA CARBONYL OXYGEN).

FT METAL 186 186 CALCIUM 1.

FT METAL 188 188 CALCIUM 1 (VIA CARBONYL OXYGEN).

FT METAL 190 190 CALCIUM 1 (VIA CARBONYL OXYGEN).

FT METAL 274 274 CALCIUM 2 (VIA CARBONYL OXYGEN).

FT METAL 276 276 CALCIUM 2 (VIA CARBONYL OXYGEN).

FT METAL 279 279 CALCIUM 2 (VIA CARBONYL OXYGEN).

FT HELIX 117 120

FT TURN 121 123

FT HELIX 125 130

FT TURN 131 131

FT TURN 135 136

FT STRAND 138 142

FT TURN 150 151

FT STRAND 156 157

FT HELIX 173 182

FT STRAND 198 201

FT TURN 207 208

FT HELIX 213 226

FT STRAND 230 233

FT HELIX 242 252

FT TURN 253 255

FT STRAND 257 258

FT TURN 262 269

FT TURN 272 275

FT STRAND 279 279

FT STRAND 282 285

FT STRAND 291 291

FT TURN 299 300

FT STRAND 303 306

FT STRAND 310 314

FT TURN 315 317

FT STRAND 318 322

FT HELIX 325 342

FT TURN 344 345

FT HELIX 348 358

FT HELIX 375 380

SQ SEQUENCE 380 AA; 539EA72771B6682C CRC64;

Query Match 10.7%; Score 240; DB 1; Length 380;

Best Local Similarity 31.1%; Pred. No. 1.1e-08;

Matches 91; Conservative 30; Mismatches 90; Indels 82; Gaps 14;

Qy 8 VKADVAQSSVGLYQCGQIVAVADTGLDTCRNDSSMHEAFRGKITALYALGRTRNANDTNG 67

Db 122 VQAPAAHNR-GLTSGVGKVAVLDTGIST-----HPDLNIRGGASFVPGSP-STQDNG 172

Qy 68 HGTHVAGSV--LNGSTNKGMAPQANLVPQSIMDSGGGLGGLPSNLQTLFQAYSAGARI 125

Db 173 HGTHVAGTIAALNNSIGVLGVAIPAELYAVKVLGASG--SGSVSSIAQGLEWAGNNGMHV 230

Qy 126 HTNSWGA-----AVNGAYTDSRNVDDYVRKNDMTILFAAGNEGNGGTISAPGTA 176

Db 231 ANLSLGSPPSATLEQAVNSA---TSRGV-----LVVAAGNSG--AGSISYPARY 276

Qy 177 KNATVIGATE--NLRPFSFGSYADNINHVAFSSRGPTKGRIPKPDVMAFGTILSARSSL 234

Db 277 ANAVAVGATDQNNNRASFQYAGL-----DIVAFGVNVQSTYFG- 316

QY 235 APDSFWANHDSKYAYMGTSMTPIVAGNVA-----OLREHFVN 275
 Db 317 -----STYASLNGTSMATPHVAGAAALVKQKNSWSNVQIRNH-LKN 357

RESULT 14
 ELVA_BACCS
 ID ELVA_BACCS STANDARD; PRT; 380 AA.
 AC P41362;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Alkaline protease precursor (EC 3.4.21.-).
 OS Bacillus clausii.
 OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=9880;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
 RX MEDLINE=93043753; PubMed=1368952;
 RA Takami H., Kobayashi T., Kobayashi M., Yamamoto M., Nakamura S.,
 Aono R., Horikoshi K.;
 RT "Molecular cloning, nucleotide sequence, and expression of the
 RT structural gene for alkaline serine protease from alkaliphilic
 RT Bacillus sp. 221.";
 RL Biosci. Biotechnol. Biochem. 56:1455-1460(1992).
 [2]
 RN SEQUENCE OF 112-129.
 RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
 RA Horikoshi K.;
 RL (In) Horikoshi K. (eds.);
 RL Microorganisms in alkaline environments, pp.187-194, VCH,
 RL Weinheim (1991).
 CC -!- COFACTOR: Binds 2 calcium ions per subunit (By similarity).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; S48754; AAC60420.1; -;
 CC EMBL; D13157; BAA02442.1; -;
 CC EMBL; A26817; CAA01836.1; -;
 CC EMBL; A22550; CAA01611.1; -;
 CC HSP; P29600; IGC1.
 CC MEROPS; S08.103; -;
 CC InterPro; IPR000209; Peptidase_S8.
 CC InterPro; IPR009020; Protease_inhib.
 CC Pfam; PF00082; Peptidase_S8; 1.
 CC PRINTS; PR00723; SUBTILISIN.
 CC PROSITE; PS00136; SUBTILASE_HIS; 1.
 CC PROSITE; PS00137; SUBTILASE_HIS; 1.
 CC PROSITE; PS00138; SUBTILASE_SER; 1.
 CC Hydrolase; Serine protease; Zymogen; Metal-binding; Calcium-binding;
 CC Signal.
 KW SIGNAL.
 FT SIGNAL 1 27 POTENTIAL.
 FT PROPEP 28 111
 FT CHAIN 112 380
 FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 326 326 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT METAL 113 113 CALCIUM 1 (BY SIMILARITY).
 FT METAL 151 151 CALCIUM 1 (BY SIMILARITY).
 FT METAL 184 184 CALCIUM 1 (VIA CARBONYL OXYGEN)
 FT METAL 186 186 (BY SIMILARITY).
 FT METAL 186 186 CALCIUM 1 (BY SIMILARITY).
 FT METAL 188 188 CALCIUM 1 (VIA CARBONYL OXYGEN)

FT METAL 190 190 (BY SIMILARITY).
 FT METAL 274 274 CALCIUM 1 (VIA CARBONYL OXYGEN)
 FT METAL 274 274 (BY SIMILARITY).
 FT METAL 276 276 CALCIUM 2 (VIA CARBONYL OXYGEN)
 FT METAL 276 276 (BY SIMILARITY).
 FT METAL 279 279 CALCIUM 2 (VIA CARBONYL OXYGEN)
 FT METAL 279 279 (BY SIMILARITY).
 SQ SEQUENCE 380 AA; 38826 MW; 5F73ABC68D5B6831 CRC64;
 Query Match 10.7%; Score 240; DB 1; Length 380;
 Best Local Similarity 31.1%; Pred. No. 1.1e-08;
 Matches 91; Conservative 30; Mismatches 90; Indels 82; Gaps 14;
 QY 8 VKADVAOSSYGLYGGQGVAVADTGLDTGRNDSMHEAFRGKITALYALGRITNNANTNG 67
 Db 122 VQAPAAHNR-CLTSGVKVAVLDTGIST-----HPDLNIRGGASFVPGEF-STQDGG 172
 QY 68 HGTHTVAGSV--LNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYSAGARI 125
 Db 173 HGTHVAGTIAALNNSIGVLGVAPSAELYAVKVLGASG--SGSVASSIAQGLEWAGNNGMHV 230
 QY 126 HTNSWGA-----AVNGAYTTDSRNVDYVRKNDMTILFAAGNGGNGGTISAFGTA 176
 Db 231 ANLSGSPSPSATLEQAVNSA---TSRGV-----LVVAASGNSG--AGSISYPARY 276
 QY 177 KNAITVGATE--NLRFPGSVYADNINHVAFSPSRGPTKGRIPKDPVMAFGTIFLSARSSL 234
 Db 277 ANAMAVGATDQNNNRASPSQYAGL-----DIVAPGVNVQSTYPG- 316
 QY 235 APDSFWANHDSKYAYMGTSMTPIVAGNVA-----OLREHFVN 275
 Db 317 -----STYASLNGTSMATPHVAGAAALVKQKNSWSNVQIRNH-LKN 357

RESULT 15
 PRTM_BACSP
 ID PRTM_BACSP STANDARD; PRT; 269 AA.
 AC Q99405;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE M-protease (EC 3.4.21.-).
 OS Bacillus sp. (strain KSM-K16).
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1409;
 RN [1]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RA Yamane T., Kani T., Hatanaka T., Suzuki A., Ashida T.,
 RA Kobayashi T., Ito S., Yamashita O.;
 RT "Structure of a new alkaline serine protease (M-protease) from
 RT Bacillus sp. KSM-K16.";
 RL Acta Crystallogr. D 51:199-206(1995).
 RN [2]
 RP SEQUENCE OF 1-23, AND CHARACTERIZATION.
 RX MEDLINE=95358832; PubMed=7632397;
 RA Kobayashi T., Hakamada Y., Adachi S., Hitomi J., Yoshimatsu T.,
 RA Koike K., Kawai S., Ito S.;
 RT "Purification and properties of an alkaline protease from
 RT alkaliphilic Bacillus sp. KSM-K16.";
 RL Appl. Microbiol. Biotechnol. 43:473-481(1995).
 CC -!- COFACTOR: Binds 2 calcium ions per subunit.
 CC -!- SUBUNIT: Monomer.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC PDB; 1MPT; 22-JUN-94.
 DR MEROPS; S08.010; -;
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_HIS; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.

KW Hydrolase; Serine protease; Metal-binding; Calcium-binding;
KW 3D-structure. 32 32 CHARGE RELAY SYSTEM.
FT ACT SITE 62 62 CHARGE RELAY SYSTEM.
FT ACT SITE 215 215 CHARGE RELAY SYSTEM.
FT METAL 2 2 CALCIUM 1.
FT METAL 40 40 CALCIUM 1.
FT METAL 73 73 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 75 75 CALCIUM 1.
FT METAL 77 77 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 79 79 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 163 163 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT METAL 165 165 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT METAL 168 168 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT METAL 168 168 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT HELIX 6 10
FT TURN 11 12
FT TURN 13 18
FT TURN 19 20
FT TURN 24 25
FT TURN 27 32
FT TURN 39 40
FT TURN 43 48
FT TURN 51 52
FT TURN 62 71
FT TURN 84 85
FT TURN 87 92
FT TURN 96 97
FT TURN 102 114
FT TURN 115 116
FT TURN 119 122
FT TURN 126 126
FT TURN 131 142
FT TURN 143 144
FT TURN 146 150
FT STRAND 161 161
FT TURN 162 164
FT TURN 166 167
FT TURN 169 174
FT TURN 176 177
FT STRAND 180 180
FT TURN 182 183
FT TURN 188 189
FT STRAND 192 195
FT STRAND 199 203
FT TURN 204 206
FT STRAND 207 211
FT HELIX 214 231
FT TURN 233 234
FT TURN 237 246
FT TURN 247 247
FT STRAND 249 249
FT HELIX 254 257
FT TURN 258 259
FT STRAND 261 261
FT HELIX 264 267
FT TURN 268 269
SQ SEQUENCE 269 AA; 26723 MW; 7A03C86D534A1D07 CRC64;

Query Match 10.6%; Score 239; DB 1; Length 269;
Best Local Similarity 31.1%; Pred. No. 8e-09;
Matches 91; Conservative 30; Mismatches 90; Indels 92; Gaps 14;
Qy 8 VKADVAQSSYGLYGGQIVAVADTGLDGRNDSSMEHAFRGKITALYALGRTNNDYNG 67
Db 11 VQAPAAHNR-GLAGSGVKVAVLDTGIST-----HPDLNIRGGASFVPEP-STQDQNG 61
Qy 68 HGTHVAGSV--LGNSTNKGMAQANLVFQSIINDSGGLGLPSNLQTLFSCAYSAGARI 125
Db 62 HGTHVAGTIAALNNSIGVLGVPASAEIYAVKVLGASG--SGSVSSIAQGLEWAGNMGHV 119
Qy 126 HTNSWGA-----AVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTA 176
Db 120 ANLSLGSFSPSATLEQAVNSA---TSRGV-----LVVAASGNSG--AGSISYPARY 165

Qy 177 KNAITVGATE--NLRPFGSYADNINHVAFSSRGPTKDGRIPDVMAPGTILSARSSL 234
Db 166 ANRMVAGTADQNNNRASFQYAGL-----DIVAFGVNVQSTYPG- 205
Qy 235 APDSFWANHDSKYAYMGGTSMATPTVAGNVA-----QUREHFVN 275
Db 206 -----STYASLNGTSMATPHVAGVAALVKQKQNPWSNVQIRNH-LKN 246

Search completed: March 10, 2004, 14:45:18
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:42:22 ; Search time 46 Seconds
(without alignments)
2976.846 Million cell updates/sec

Title: US-09-985-689a-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAYNVPGPQTFSLAIVN 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL 25:*
- 1: sp archea:*
 - 2: sp bacteria:*
 - 3: sp fungi:*
 - 4: sp human:*
 - 5: sp invertebrate:*
 - 6: sp mammal:*
 - 7: sp mhc:*
 - 8: sp organelle:*
 - 9: sp phage:*
 - 10: sp plant:*
 - 11: sp rodent:*
 - 12: sp virus:*
 - 13: sp vertebrate:*
 - 14: sp unclassified:*
 - 15: sp virus:*
 - 16: sp bacteriap:*
 - 17: sp archaep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2247	100.0	640	2	Q93UV9	Q93UV9 bacillus sp
2	2183	97.2	639	2	Q9AQR3	Q9AQR3 bacillus sp
3	2143	95.4	434	2	Q9AQR0	Q9AQR0 bacillus sp
4	1998.5	88.9	433	2	Q9AQR1	Q9AQR1 bacillus sp
5	1994.5	88.8	433	2	Q9AQR4	Q9AQR4 bacillus sp
6	1987.5	88.5	433	2	Q9AQR2	Q9AQR2 bacillus sp
7	511.5	22.8	1825	5	Q8T9W1	Q8T9W1 dictyosteli
8	447	19.9	1702	5	Q8GTN7	Q8GTN7 dictyosteli
9	414	18.4	654	17	Q8UOC9	Q8UOC9 pyrococcus
10	406.5	18.1	561	16	Q8RBZ2	Q8RBZ2 thermoaer
11	398	17.7	1239	16	Q9FBZ4	Q9FBZ4 streptomyc
12	381	17.0	430	16	Q8ENV1	Q8ENV1 oceanobacil
13	376	16.7	1253	16	Q9FCO6	Q9FCO6 streptomyc
14	366	16.3	1102	2	P95684	P95684 streptomyc
15	363.5	16.2	1208	16	Q82B14	Q82B14 streptomyc
16	349	15.5	1245	16	Q9RL54	Q9RL54 streptomyc

ALIGNMENTS

RESULT 1

Q93UV9 ID Q93UV9 PRELIMINARY; PRT; 640 AA.
AC Q93UV9; 19, Created)
DT 01-DEC-2001 (TREMREL. 19, Created)
DT 01-OCT-2002 (TREMREL. 22, Last sequence update)
DT 01-JUN-2003 (TREMREL. 24, Last annotation update)
DE Protease.
GN PROF.
OS Bacillus sp. KSM-KP43.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=109322;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KP43;
RA Itoh S., Saeki K.;
RT "new protease.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051423; BAB55674.2; -
DR GO; GO:0004289; P:protease activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 640 AA; 67991 MW; 48BAF7E9D592C15 CRC64;

Query Match 100.0%; Score 2247; DB 2; Length 640;
Best Local Similarity 100.0%; Pred. No. 3.8e-122;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGGQGVAVADTGLDTRNDSSMHEAFRGKITALVALGRIN 60
DB 207 NDVARGIVKADVAQSSYGLYGGQGVAVADTGLDTRNDSSMHEAFRGKITALVALGRIN 266
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFOSIMDSGGGLGFLPSNLQTLFSQAYS 120


```

Db 267 NANTDNGHGTAVAGSVLNGSTNKGMAPOANLVFQSIMDSSGGGLGSLPSNLQTLFSQAYS 326
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 327 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
QY 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAPGTFILSARSLAPDSSF 240
Db 387 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAPGTFILSARSLAPDSSF 446
QY 241 WANHDSKYAYMGTSWATPIVAGNVAQLREHFVKNGITPKPSLLKAALIAAGADIGLY 300
Db 447 WANHDSKYAYMGTSWATPIVAGNVAQLREHFVKNGITPKPSLLKAALIAAGADIGLY 506
QY 301 PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 360
Db 507 PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 566
QY 361 SVTLVNDLVLITAPNGTQVGNDFTSYNDWNGRNNVENFINAPQSGTITIEVQAYN 420
Db 567 SVTLVNDLVLITAPNGTQVGNDFTSYNDWNGRNNVENFINAPQSGTITIEVQAYN 626
QY 421 VPVGPQTFSLAIVN 434
Db 627 VPVGPQTFSLAIVN 640

```

RESULT 2

```

Q9AQR3 ID Q9AQR3 PRELIMINARY; PRT; 639 AA.
AC Q9AQR3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Protease.
GN PROA.
OS Bacillus sp. 9860.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=133778;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9860;
RX MEDLINE=20568675; PubMed=11118284;
RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus sp.: enzymatic properties, sequences, and
RT evolutionary relationships."
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
DR EMBL; AB046403; BAB21266.2; -.
DR HSSP; P00782; 1SUP.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 639 AA; 68185 MW; 316AF6FDBE4FF54 CRC64;

```

Query Match 97.3%; Score 2183; DB 2; Length 639;
 Best Local Similarity 96.3%; Pred. No. 1.9e-118;
 Matches 418; Conservative 13; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 NDVARGIVKADVQSSVGLYGQGIIVAVADTGLDGRNDSMHEAPRGKITALYALGRN 60
Db 206 NDVARGIVKADVQSSVGLYGQGIIVAVADTGLDGRNDSMHEAPRGKITALYALGRN 265
QY 61 NANTDNGHGTAVAGSVLNGSTNKGMAPOANLVFQSIMDSSGGGLGSLPSNLQTLFSQAYS 120

```

```

Db 266 NANTDNGHGTAVAGSVLNGSTNKGMAPOANLVFQSIMDSSGGGLGSLPSNLQTLFSQAFS 325
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 326 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 385
QY 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAPGTFILSARSLAPDSSF 240
Db 386 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAPGTFILSARSLAPDSSF 445
QY 241 WANHDSKYAYMGTSWATPIVAGNVAQLREHFVKNGITPKPSLLKAALIAAGADIGLY 300
Db 446 WANHDSKYAYMGTSWATPIVAGNVAQLREHFVKNGITPKPSLLKAALIAAGADIGLY 505
QY 301 PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 360
Db 506 PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 565
QY 361 SVTLVNDLVLITAPNGTQVGNDFTSYNDWNGRNNVENFINAPQSGTITIEVQAYN 420
Db 566 SVTLVNDLVLITAPNGTQVGNDFTSYNDWNGRNNVENFINAPQSGTITIEVQAYN 625
QY 421 VPVGPQTFSLAIVN 434
Db 626 VPVGPQTFSLAIVN 639

```

RESULT 3

```

Q9AQR0 ID Q9AQR0 PRELIMINARY; PRT; 434 AA.
AC Q9AQR0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Protease (Fragment).
GN PROE.
OS Bacillus sp. NV1.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=133781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NV1;
RX MEDLINE=20568675; PubMed=11118284;
RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus sp.: enzymatic properties, sequences, and
RT evolutionary relationships."
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
DR EMBL; AB046406; BAB21269.1; -.
DR HSSP; P00782; 1SUP.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydroxylase; Protease; Serine protease.
FT NON_TER 434
FT NON_TER 434
SQ SEQUENCE 434 AA; 45294 MW; 83517EDDB74125D2 CRC64;

```

Query Match 95.4%; Score 2143; DB 2; Length 434;
 Best Local Similarity 93.5%; Pred. No. 2.4e-116;
 Matches 406; Conservative 19; Mismatches 9; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVQSSVGLYGQGIIVAVADTGLDGRNDSMHEAPRGKITALYALGRN 60

Db 1 NDVARGIKADVAQSSVGLYGGQVAVADTGLDTGRNDSMHEAFRGKITAIYALGRTN 60
 QY 61 NNDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
 Db 61 NNDPNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSVMDNSGGLGGLPSNVSTLFSQAYS 120
 QY 121 AGARIHTNSGAAVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
 Db 121 AGARIHTNSGAPVNGAYTTDSNRVDDYVRKNDMAVLFAAGNEGPNNGGTISAPGTAKNAI 180
 QY 181 TVGATENLRPSFSGYADININHAQFSSRGPTKGRIPKDPVMACTFILSARSSLAPDSF 240
 Db 181 TVGATENLRPSFSGYADININHAQFSSRGPTKGRIPKDPVMACTFILSARSSLAPDSF 240
 QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY 300
 Db 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY 300
 QY 301 PNGNGGGRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLWSDAPASTTA 360
 Db 301 PNGNGGGRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLWSDAPASTTA 360
 QY 361 SVTLVNDLDLVTAPNGTQVGNDFTSFYNDNWDGRNNVENVFINAPQSGTYTIEVQAYN 420
 Db 361 SVTLVNDLDLVTAPNGTQVGNDFTSFYNDNWDGRNNVENVFINAPQSGTYTIEVQAYN 420
 QY 421 VPVGPQTFSLAIVN 434
 Db 421 VPQGPQAFSLAIVN 434

RESULT 4

Q9AQR1 ID Q9AQR1 PRELIMINARY; PRT; 433 AA.
 AC Q9AQR1;
 DT 01-JUN-2001 (TremBLrel. 17, Created)
 DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
 DE Protease (Fragment).
 GN PROD.
 OS Bacillus sp. SD521.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=133780;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SD521;
 RX MEDLINE=20568675; PubMed=11118284;
 RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
 RA Horikoshi K.;
 RT "Novel oxidatively stable subtilisin-like serine proteases from
 RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
 RT evolutionary relationships.";
 RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
 DR EMBL; AB046405; BAB2128.1; -;
 DR HSSP; Q45670; 1DBI.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004289; F:subtilase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000209; Peptidase_S8.
 DR InterPro; IPR007280; PPC.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF04151; PPC; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00137; SUBTILASE HIS; 1.
 DR PROSITE; PS00138; SUBTILASE SER; 1.
 KW Hydrolase; Protease; Serine protease.
 FT NON_TER 1
 FT NON_TER 433
 SQ SEQUENCE 433 AA; 45576 MW; 98A2DF18FE65DDC CRC64;

Query Match 88.9%; Score 1998.5; DB 2; Length 433;
 Best Local Similarity 87.8%; Pred. No. 5.5e-108;

Matches 381; Conservative 28; Mismatches 24; Indels 1; Gaps 1;
 QY 1 NDVARGIKADVAQSSVGLYGGQVAVADTGLDTGRNDSMHEAFRGKITAIYALGRTN 60
 Db 1 NDVARGIKADVAQSSVGLYGGQVAVADTGLDTGRNDSMHEAFRGKITAIYALGRTN 60
 QY 61 NNDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
 Db 61 NNDPNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSVMDNSGGLGGLPSNVSTLFSQAYS 119
 QY 121 AGARIHTNSGAAVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
 Db 120 AGARIHTNSGAPVNGAYTANSRQVDEVNRNDMTVLFAAGNEGPNNGGTISAPGTAKNAI 179
 QY 181 TVGATENLRPSFSGYADININHAQFSSRGPTKGRIPKDPVMACTFILSARSSLAPDSF 240
 Db 180 TVGATENLRPSFSGYADININHAQFSSRGPTKGRIPKDPVMACTFILSARSSLAPDSF 239
 QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY 300
 Db 240 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY 299
 QY 301 PNGNGGGRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLWSDAPASTTA 360
 Db 300 PNGNGGGRVTLDKSLNVAYVNEATATATGOKATYSFOAQAGKPLKISLWSDAPASTTA 359
 QY 361 SVTLVNDLDLVTAPNGTQVGNDFTSFYNDNWDGRNNVENVFINAPQSGTYTIEVQAYN 420
 Db 360 SVTLVNDLDLVTAPNGTQVGNDFTSFYNDNWDGRNNVENVFINAPQSGTYTIEVQAYN 419
 QY 421 VPVGPQTFSLAIVN 434
 Db 420 VPQGPQAFSLAIVH 433

RESULT 5

Q9AQR4 ID Q9AQR4 PRELIMINARY; PRT; 433 AA.
 AC Q9AQR4;
 DT 01-JUN-2001 (TremBLrel. 17, Created)
 DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
 DE Protease (Fragment).
 GN PROD.
 OS Bacillus sp. D6.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=127889;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D6;
 RX MEDLINE=20568675; PubMed=11118284;
 RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
 RA Horikoshi K.;
 RT "Novel oxidatively stable subtilisin-like serine proteases from
 RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
 RT evolutionary relationships.";
 RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
 DR EMBL; AB046402; BAB21265.1; -;
 DR HSSP; Q45670; 1DBI.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004289; F:subtilase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000209; Peptidase_S8.
 DR InterPro; IPR007280; PPC.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF04151; PPC; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00137; SUBTILASE HIS; 1.
 DR PROSITE; PS00138; SUBTILASE SER; 1.
 KW Hydrolase; Protease; Serine protease.
 FT NON_TER 1
 FT NON_TER 433

SQ SEQUENCE 433 AA; 45636 MW; 52087BOA2516107P CRC64;

Query Match 88.8%; Score 1994.5; DB 2; Length 433;
 Best Local Similarity 87.6%; Pred. No. 9.4e-108;
 Matches 380; Conservative 28; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITALVALGRTN 60
 DB 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITALVALGRTN 60

QY 61 NANDTNGHGHVAGSVLGN-ALNKGWAPQANLVFQSIMDSGGGLGGLPSNLTLFSQAWN 120
 DB 61 NANDTNGHGHVAGSVLGN-ALNKGWAPQANLVFQSIMDSGGGLGGLPSNLTLFSQAWN 119

QY 121 AGARIHTNSWGAVNGAYTTDSRNVDVYRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 120 AGARIHTNSWGAVNGAYTTDSRNVDVYRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 179

QY 181 TVGATENLRPSFGSIADNPENHIAQFSSRGATRDGRIKPDVTAPGTIFLSARSSLPDSSF 240
 DB 180 TVGATENLRPSFGSIADNPENHIAQFSSRGATRDGRIKPDVTAPGTIFLSARSSLPDSSF 239

QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGADIGLY 300
 DB 240 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGADIGLY 299

QY 301 PNGNQGWGRVTLKSLNVAVYVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 360
 DB 300 PNGNQGWGRVTLKSLNVAVYVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 359

QY 361 SVTLVNDLVLITAPNGTQYVGNDFPSYNDNDGNNVNFVINAPOSCTTIEVOAYN 420
 DB 360 SVTLVNDLVLITAPNGTQYVGNDFPSYNDNDGNNVNFVINAPOSCTTIEVOAYN 419

QY 421 VPVGPQTFSLAIYN 434
 DB 420 VPVGPQTFSLAIYH 433

RESULT 6

Q9AQR2 PRELIMINARY; PRT; 433 AA.

AC Q9AQR2;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Protease (Fragment).
 GN PROC.
 OS Bacillus sp. Y.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 CX NCBI_TaxID=133779;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Y;
 RX MEDLINE=20568675; PubMed=11118284;
 RA Saeki K.; Okuda M.; Hatada Y.; Kobayashi T.; Ito S.; Takami H.;
 RA Horikoshi K.;
 RT "Novel oxidatively stable subtilisin-like serine proteases from
 RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
 RT evolutionary relationships.";
 RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
 DR EMBL; AB046404; BAB21267.1; -;
 DR HSP; Q45670; 1DB1.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004289; F:subtilase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.

DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 FT NON_TER 433 1
 SQ SEQUENCE 433 AA; 45587 MW; B81291A803C75AE CRC64;

Query Match 88.5%; Score 1987.5; DB 2; Length 433;
 Best Local Similarity 87.3%; Pred. No. 2.4e-107;
 Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITALVALGRTN 60
 DB 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITALVALGRTN 60

QY 61 NANDTNGHGHVAGSVLGN-ALNKGWAPQANLVFQSIMDSGGGLGGLPSNLTLFSQAWN 120
 DB 61 NANDTNGHGHVAGSVLGN-ALNKGWAPQANLVFQSIMDSGGGLGGLPSNLTLFSQAWN 119

QY 121 AGARIHTNSWGAVNGAYTTDSRNVDVYRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 120 AGARIHTNSWGAVNGAYTTDSRNVDVYRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 179

QY 181 TVGATENLRPSFGSIADNPENHIAQFSSRGATRDGRIKPDVTAPGTIFLSARSSLPDSSF 240
 DB 180 TVGATENLRPSFGSIADNPENHIAQFSSRGATRDGRIKPDVTAPGTIFLSARSSLPDSSF 239

QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGADIGLY 300
 DB 240 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGADIGLY 299

QY 301 PNGNQGWGRVTLKSLNVAVYVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 360
 DB 300 PNGNQGWGRVTLKSLNVAVYVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 359

QY 361 SVTLVNDLVLITAPNGTQYVGNDFPSYNDNDGNNVNFVINAPOSCTTIEVOAYN 420
 DB 360 SVTLVNDLVLITAPNGTQYVGNDFPSYNDNDGNNVNFVINAPOSCTTIEVOAYN 419

QY 421 VPVGPQTFSLAIYN 434
 DB 420 VPVGPQTFSLAIYH 433

RESULT 7

Q8T9W1 PRELIMINARY; PRT; 1825 AA.

AC Q8T9W1;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Serine protease/ABC transporter TagD.
 GN TAGD.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 CX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ax4;
 RA Anjard C.; Loomis W.F.;
 RT "Evolution of the ABC transporters of Dictyostelium";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; AF466309; AAL74253.1; -;
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.
 DR GO; GO:0001656; F:nucleotide binding; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004289; F:subtilase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR001140; ABC_TM_transp.


```
RESULT 9
Q8UBC9 PRELIMINARY; PRT; 654 AA.
AC Q8UBC9
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Alkaline serine protease.
GN PF1670.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AAL81794.1; -.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PRO0723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 654 AA; 70230 MW; 1CB145A5F505DB34 CRC64;

Query Match 18.4%; Score 414; DB 17; Length 654;
Best Local Similarity 29.8%; Pred. No. 5.4e-16;
Matches 136; Conservative 59; Mismatches 147; Indels 144; Gaps 19;

QY 18 GLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRNNAN-----DTNGHGH 71
DB 154 GYDGGITGIIDTGID-----ASHPDLGQKV-----IGWDFVNGRSYFDDHGHGH 202

QY 72 VAGSVLNGSTN-----KGMAPQNLVFGSIM--DSGGGLGLPLSNLTLSQAYSAGARI 125
DB 203 VASIAAGTGAASNGYKGMAPGAKLAGIKVLGADGSGSITTIKGVAVDNKDKYGIKV 262

QY 126 HTNSMGA-----AVNGAYTDSRVDDYVRKNDMTILFAAGNEGNGGTISAP 173
DB 263 INLSGSSSSDGTALSOAVNAWDA-----GLVVVAAGNSGPNKVTIGSP 310

QY 174 GTAKNAITVATENLRPSFGSVADNINHVAQFSRGPDKGRIPKDPVMAPTFILSARSS 233
DB 311 AAASKVITVGA-----VDKYDVTSPSRGPTADGRKLPDEVVAPGNWIIAARAS 359

QY 234 LAPDSFWANHDSKYAYMGTSNATPIVAGNVAQIREHFVKRGITPK--PSLLKAALIA 291
DB 360 -----GTSMGQFINDYITAPGTSNATPHVAGIALLLQ-----AHPSTPKVKTALIE 409

QY 292 GA-----ADIGLYPNGNGQGRVTLDSINVAYNNESSLSLSOKA-----TYSFT 338
DB 410 TADIVKPEIDAIYGA-----GRVNAYKAIN--YDNYAKLVFTGVVANKGSQTHQFV 460

QY 339 ATAGKPLKLSLVWSDAPASTTASVTLVNDLVLITAPNGTQVYVNDFTSPYNDNWDGRNN 396
DB 461 ISGASEVTATLYWDNAN-----SLLDLYLDPNGNQ--VDYSYATAY-----G 501

QY 399 VENVFNAQPSGTYTIEVOAYNVVGPQTFSIAIYN 434
DB 502 FEKVGYNNTDGTWIKVYSYS---GSANYQVDVVS 534
```

RESULT 10
Q8RBJ2

```
Q8RBJ2 PRELIMINARY; PRT; 561 AA.
Q8RBJ2
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Subtilisin-like serine proteases.
GN APRE2 OR TTE0824.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AEO13049; AM24081.1; -.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00728; PPC.
DR PRINTS; PRO0728; PPC.
DR InterPro; IPR009020; Protease_inhib.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PRO0723; SUBTILISIN.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 561 AA; 59968 MW; BA9C5C2F7083A18 CRC64;

Query Match 18.1%; Score 406.5; DB 16; Length 561;
Best Local Similarity 31.3%; Pred. No. 1.2e-15;
Matches 144; Conservative 62; Mismatches 135; Indels 119; Gaps 22;

QY 6 GIVKADVQSSYGLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRNNAN-- 63
DB 155 GITK--ARSDFGVTGKNITAIIDTGDGNHVDLS-----GGKI-----IGWKFINK 201

QY 64 -----DTNGHGHVAGSVLNGSTN---KGMAPQNLVFGSIMDSGGG-----LGLPSN 110
DB 202 TTYDDNGHGHVASTAAGTGAGNSFYKGVAPDALLVIGIKVLDANGSGMSVTFAGIDWA 261

QY 111 LQ-----TLFSQAYSAGARIHNSWGAANGAYTTDSRVDDYVRKNDMTILF 158
DB 262 VQNKDVVGIKVINLSLGTSTSDG-----TDSLSLVN-----RAVD-----SGIVVV 305

QY 159 AAGNEGNGGTISAPGTAKNAITVATENLRPSFGSVADNINHVAQFSRGPDKGRIPK 218
DB 306 AAGNSGPAKTIISGPAEAKAITVAAMADV-----GELGNL---ASFSGPTADGRIPK 358

QY 219 DVMAPTFILSARSSLAPDSSFWANHDSKYAYMGTSNATPIVAGNVAQIREHFVKRGI 278
DB 359 DIAAPGNITAAK-----ANSVNGVVTYSGTSMATPFVAGTVALMLN---ANPNL 405

QY 279 TPSPSLLKAALIAAGADIGLYNGNGQGRVTLDSINVA-----YVNES 324
DB 406 TPNDAA--KNITNGTAKSWGPPSKNDYAGRDLGYEIRVAGNFRGNNDVPHNYTI--S 461

QY 325 SLSLSOKA--TYSFTAT--AGKPLKISLV---WSDAPASTTASVTLVNDLVLITAPNGTQ 379
DB 462 GYLPFSKYSDTWTFTNATISYPIAITLIIPDMANYNP-----DPIYLYDPSGTL 511

QY 380 YVGNDFTSFYNDNWDGRNNVFVFNAPSGTYTIEVOAY 419
```


Db 173 EYDD-NHGTGHCAGDAAGNALSQYQGPAPDANLVGVKYNKGTGS-GSLSTVIEGI- 229
QY 116 SQAYSAGARIHTNSGAAVNGAYTTDSRVDDYVR-----KNDMTILFAAGNEGPNCGT 169
Db 230 DNCIQNCIKINILSLSL-GSDATEPABGDPVNAVETANDNGMVCVAAGNSGPGDKT 288
QY 170 ISAPGTAKNAITVGATENLRPFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTFLS 229
Db 289 VSGPGISPKVITVGAADDNNTAERS-----DDSAEFSRGPTIDGLTKPNTLLTPGVDIVS 344
QY 230 ABS--SLAPDSSFWANHDSKYAYMGTSNATIVAGNVAQLBEHFVKRGITPKPSLLKA 287
Db 345 LRAPGSFIDTKNKSARVNGSYLSLGSNTATPICAGIAVQLLQ---SDSSLT--PNQVKE 399
QY 288 ALIAGAADIGLGPNGNQGWRVTLDKSLNVAVNE 323
Db 400 KLMEACQDLIGS-PN-VQAGYL---NPAANLINE 430

RESULT 13

Q9FC06 PRELIMINARY; PRT; 1253 AA.
AC Q9FC06; MEDLINE=21996410; PubMed=12000953;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative secreted peptidase.
GN SC07176 OR SC8A11.04C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]_Strain=1902;
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Denapalte D., Eichner A., Cullum J.,
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Hopwood D.A.;
RL "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RN Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RP "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AJ239130; CAC01576.1; -.
DR HSSP; Q99405; IMPT.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Complete proteome.
SQ SEQUENCE 1253 AA; 130971 MW; AA69B417EFEEB89 CRC64;

Query Match 16.7%; Score 376; DB 16; Length 1253;

Best Local Similarity 27.6%; Pred. No. 2e-13;
Matches 142; Conservative 58; Mismatches 145; Indels 170; Gaps 22;

QY 16 SYGLYGQGIYAVADTGLDTGRNDSMHEAFRGKITALYALGRTNNANDTNGHGTAVGS 75
Db 233 SGGNTGEGVAVLDTGVDAG-----HPDFAGRIATAATSFVPDQDVTDRNGHGTAVST 286
QY 76 VLNGSTN-----KGMAPQANLVFQSTMDSGG-----GLGG 106
Db 287 VAGTGAASGGVEKGVAPGASLHIGKVLDSGSGQSDSWLAGMEWAVRDQHAQIVMSLGD 346
QY 107 LPSNLQTLFSQAYSAGARIHTNSWGAANVAGAYTTDSRVDDYVRKNDMTILFAAGNEGN 166
Db 347 SPTDGTDLSEAV-----NWLSAETGA-----LFVVAAGNSGPE 380
QY 167 GTTISAPGTAKNAITVGATENLRPFGSYADNINHVAQFSSRGCP-TKDGRIKPDVMAFGT 225
Db 381 AYTGTGTPAAADAALITGAVNG--PGRG-----VDQLADFSRGRVGVGDNAVKPDLTAPGV 433
QY 226 FILSARSSLAPDSSFWANHDSKYAYMGTSNATIVAGNVAQLBEHFVKRGITPKPSLL 285
Db 434 GVLAARSRYAPEG-----EGAYQSLSGTSMATPHVAGAAALAAEHPDWTG-----QRL 482
QY 286 KAALIAGAADIGLGPNGNQ-----GMGRVTLKSLNVAVYNESSLSSTQ---KATY 335
Db 483 KEALVGTGTA-----GTQRFSPFDAGSGRV-----DVAARVSTLLASGDAFAQAHY 528
QY 336 -----SFTATAGKPLKISLVMSDA-----PASTTASVTLVND 367
Db 529 PYTPGQTVRRDVTYNSGPAPVALDLALSPAEPLGFLTISEAQVTVPAHGHTASVGVITH 588
QY 368 LD-----LVITAPNGTQYGVNDFTSPYNDNWDGNNVNFVINA-----PQSG 410
Db 589 LAAEDNGAVATRLVASGADGAVLA---RTPGVNKEGRR--ATLALTAKDHDKPLSG 642
QY 411 T-----YTVIEVQA-YNVVPGPQTFSL 430
Db 643 TVILKQVETNAPKRVSVSDASGELDLRLSESTYSV 677

RESULT 14

P95684 PRELIMINARY; PRT; 1102 AA.
ID P95684
AC P95684; MEDLINE=97144528; PubMed=8990295;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Subtilisin-like protease.
OS Streptomyces albobogiscolus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S-3253;
RX MEDLINE=97144528; PubMed=8990295;
RA Suzuki M., Taguchi S., Yamada S., Kojima S., Miura K., Monose H.;
RL "A novel member of the subtilisin-like protease family from
Streptomyces albobogiscolus.";
RL J. Bacteriol. 179:430-438(1997).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.

```

DR EMBL; D83672; BAAL2040.1; -.
DR HSSP; P00782; 2SET.
DR MEROPS; S09.069; -.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002860; GH_BNR.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF02012; BNR; 2.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PROSITE; PRO0723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW hydrolase; Protease; Serine protease.
SQ SEQUENCE 1102 AA; 114128 MW; F9E4AD2590FE59E CRC64;

Query Match          16.3%; Score 366; DS 2; Length 1102;
Best Local Similarity 31.6%; Pred. No. 6.4e-13;
Matches 148; Conservative 55; Mismatches 175; Indels 90; Gaps 20;

QY 3 VAR-----GIVADVAQS-----SYGLYGGQIVAVADTGLDTGRNDSMHEAFRG 48
DB 183 VARVWLDGVRKASLDTSVGGITGTPEKAWEGYDGKGVKIAVLDTGVD-----ATHPDLKG 236
QY 49 KITALVALGRTNNANDTNGHGVAGSVLNGS----TNKGMAPOANLVFQISMDSGGL 104
DB 237 QVTSKNFTSAPTTGCDVVGHGTHVASAAGTGAQSKGYKGVAPGAKILNGVLDDAG-- 294
QY 105 GGLPNLQTLFQAYSAGARIHTNSWGAAVNGYTTDSRNVDYVRK--NDWTILF--AA 160
DB 295 FGDDSGIILAGMEWAAAQCADIVNNSLG----GMDTPETDPLEAAVDKLSAERKILFAAA 350
QY 161 GNEGPNGGTISAPGTAKNAITVGATENLRSPFSYADNINHVAQSSRGP-TKGRIKPD 219
DB 351 GNEGQ--SIQSPGSAUSALTVGA-----VDDKKLADFSSGTGPRUGDGNVAKPD 397
QY 220 VNAPFTFILSARSLAPDSSFWANHDSKYAMGTSVATPIVAGNVAQLREHFVKRGIT 279
DB 398 LTAPGVVDITAAAKGNDIAKEVGEKPGAGYWTISGTSWATPHVAGAAALLKQOHPE---- 452
QY 280 KPESILKALTAGAADICLG-YPNGNCGWRVILDKLSLVAVYNSSSL----- 328
DB 453 WKYAEKGLASTKD--GKYTFEGSGRGVQDKAITQTVIABPVSLSFGVQWQHPAD 509
QY 329 ---TSQKATYSFTATAGKPLKISLVMSD-----APAS--TTASVILVNDLIVITAP-NG 377
DB 510 DAPVTYKLTIRNLGCTEDVTLKLTATGPKGKAAPAGFTLGSATL-----TVPANG 561
QY 378 TQYVGNDFTSVNDNWDGRNNVENVFINAQSS-----GTYTIEVQAINV 421
DB 562 TASVDVTADTRLGGAVDGTYSAYYVATGAGQSVRTAAAVEREVESTNV 609

RESULT 15
Q82B14 PRELIMINARY; PRT; 1208 AA.
ID Q82B14
AC Q82B14;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative peptidase.
OS SAV5721.
GN Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horiikawa H., Nakazawa H., Osone T.,

```

Search completed: March 10, 2004, 14:46:17
Job time : 49 secs

RA	Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT	"Genome sequence of an industrial microorganism Streptomyces
RT	avermittilis: deducing the ability of producing secondary
RT	metabolites.";
RT	Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RL	[2]
RN	SEQUENCE FROM N.A.
RP	STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RC	MEDLINE=22608306; PubMed=12692562;
RX	Ireda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA	Sakaki Y., Hattori M., Omura S.;
RT	"Complete genome sequence and comparative analysis of the industrial
RT	microorganism Streptomyces avermiltis.";
RT	Nat. Biotechnol. 21:526-531(2003).
DR	EWEL; AP005044; BAC73433.1; --
DR	GO; GO:0008233; F:peptidase activity; IEA.
DR	GO; GO:0004289; F:subtilase activity; IEA.
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR	InterPro; IPR003137; PA.
DR	InterPro; IPR000209; Peptidase_s8.
DR	Pfam; PF02225; PA; 1.
DR	Pfam; PF00082; Peptidase_S8; 1.
DR	PRINTS; PR00723; SUBTILISIN.
DR	PROSITE; PS00136; SUBTILASE_ASP; 1.
DR	PROSITE; PS00137; SUBTILASE_HIS; 1.
DR	PROSITE; PS00138; SUBTILASE_SER; 1.
KW	Complete proteome.
SO	SEQUENCE 1208 AA; 125548 MW; E650B5E3AB5312B CRC64;

Query Match	16.2%;	Score	363.5;	DB	16;	Length	1208;		
Best Local Similarity	32.0%;	Pred. No. Ie-12;							
Matches	128;	Conservative	54;	Mismatches	145;	Indels	73;	Gaps	18;
Qy	8	VKADVAQSS-----YGLYGCQIQIVAVADTGLTGRNDSMSHEAFGKITALYALG	57						
Db	186	VEADMAESNAQIGTRAAWDAGLTGCVTVAVLDTGYDT-----THPDLAGRVSRKSF	239						
Qy	58	RTNANDNTGHGPHVAGSVLNGGS-----TNKMAPOANLVFQSIM-DSGGGLGGLPSNLQ	112						
Db	240	DGEVADRNGHGHVTVTGGSGAASDGTGRTGVAPGATLAVGKVLSDQAG-----SESQ	294						
Qy	113	TLFSQAYSA---GARHTHSWGA--AVNGAYTTD--SRNVDDYVRKNDMTILFAAGNEGP	165						
Db	295	IITAGMWAARDVRAIRIVVMSLGLSTESADG---TDPMAEAVDTLSEETGALFVVAAGNTGA	351						
Qy	166	NGGTISAPGTAKNAITVGATENLRPSFGSYADININVAQFSSRGPTK--DGRIKPDVYMAPG	224						
Db	352	PSSIGSPGAADSALTGA-----VDSSDRAAYFTSAGPRHGDNALKPDLLAAPG	399						
Qy	225	TFILSRSSLAPDSSFWANHDSKIAYNGGTSMATPIVAGNVAQLRHFVKVNRGITPKPSL	284						
Db	400	VDIARAARSOAPGTGY-----YTSMSGTSMATPHVAGVGAULLABQHPDWTGARLKDAL	452						
Qy	285	LKAA--LIAGAADIGLGPNGNGWG-RVTLDKSLNAVY-----VNESSSLSSTSQ	331						
Db	453	MSTSEQLDASVYLGAGRVSPDPAVARVATQSDLGFEHWEHPHDADRPVTKVTVTYNSSS	512						
Qy	332	KATYSFT-ATAGKPLKISLVMSDA---PASTTASVTLVND	367						
Db	513	DTTVELISLAVRGAPAGVATLADTALTVPAHGTAATVTTGD	552						

Search completed: March 10, 2004, 14:46:17
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 22:30:03 ; Search time 3605 Seconds
(without alignments)

5217.996 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSGLY.....EVQAYNPVGPQTSLAIV 434

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh

-Q=/cgn2 1/USPTO.spool/US0985689/runat 10032004 112903 19718/app query.fasta 1.583

-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0

-UNITS=bits -START=1 -END=-1 -MATRIX=bleaum2 -TRANS=human40.cdi -LIST=45

-DOALIGN=200 -THRESH=SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL

-OUTFMT=ptc -NORM=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US0985689@cgn 1 1 2372 @runat 10032004 112903 19718 -NCPU=6 -ICPU=3

-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG

-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2247	100.0	1305	6	AX839476	Sequence
2	2247	100.0	1923	1	AB051423	Bacillus
3	2247	100.0	1923	6	AR368117	Sequence
4	2242	99.8	1923	1	AB084155	Bacillus
5	2242	99.8	1923	6	AR368118	Sequence
6	2183	97.2	1920	1	AB046403	Bacillus
7	2183	97.2	1920	6	AR368116	Sequence
8	2143	95.4	1302	1	AB046406	Bacillus
9	2125.5	94.6	3003	6	AR069954	Sequence
10	2125.5	94.6	3003	6	BD062155	Nucleic a
11	1998.5	88.9	1299	1	AB046405	Bacillus
12	1994.5	88.8	1299	1	AB046402	Bacillus
13	1987.5	88.5	1299	1	AB046404	Bacillus
14	1986.5	88.4	2218	6	E03808	E03808 DNA encodin
15	632.5	28.1	6006	1	AF268611	Unculture
16	580	25.8	30205	1	AS017218	Geobacter
17	523.5	23.3	5890	3	DDU60086	U04086 Dictyosteli
18	511.5	22.8	5772	3	AF466309	AF466309 Dictyoste
19	497	22.1	6115	3	DDU20432	U20432 Dictyosteli
20	452.5	20.1	1977	6	AR201152	Sequence
21	447	19.9	5109	3	AF263455	AF263455 Dictyoste
22	425.5	18.9	12928	2	AC096673	Trypanoso
23	416.5	18.5	1236	6	AR201146	Sequence
24	416.5	18.5	1962	6	AR201155	Sequence
25	416.5	18.5	12452	1	AB010265	AB010265 Pyrococcu
26	406.5	18.1	14376	1	AB013049	AB013049 Thermoana
27	398	17.7	303450	1	SC0939130	AL939130 Streptomy
28	381	17.0	299850	1	AP004601	AP004601 Oceanobac
29	366	16.3	3348	1	D33672	D33672 Streptomyce
30	363.5	16.2	30425	1	AP005044	AP005044 Streptomy
31	351	15.6	2539	6	AR203322	Sequence
32	349	15.5	291000	1	SC0939105	AL939105 Streptomy
33	347.5	15.5	292550	1	AP001513	AP001513 Bacillus
34	346.5	15.4	302300	1	AP005034	AP005034 Streptomy
35	346	15.4	2809	6	AR202321	Sequence
36	346	15.4	3900	1	AB007809	AB007809 Streptomy
37	338	15.0	1239	1	AP305633	Thermoana
38	338	15.0	1239	1	AY028704	AY028704 Thermoana
39	338	15.0	135638	1	AF484556	AF484556 Streptomy
40	337	15.0	12295	1	AB013026	AB013026 Thermoana
41	332.5	14.8	1329	6	AX433519	AX433519 Sequence
42	331	14.7	207829	1	BSU00010	Z99113 Bacillus su
43	328.5	14.6	300800	1	SC0939112	AL939112 Streptomy
44	327.5	14.6	301150	1	AP004602	AP004602 Oceanobac
45	324.5	14.4	6854	1	BSU392330	U39230 Bacillus sp

ALIGNMENTS

RESULT 1

AX839476
LOCUS AX839476 1305 bp DNA linear PAT 15-DEC-2003
DEFINITION Sequence 2 from Patent EP1347044.
ACCESSION AX839476
VERSION AX839476.1 GI:39922766
KEYWORDS Bacillus sp. KSM-KP43
SOURCE Bacillus sp. KSM-KP43
ORGANISM Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 Okuda, M.K., Sato, T.K., Saito, K.K., Sumitomo, N.K., Izawa, Y.K.,
AUTHORS Saeki, K.K., Kobayashi, T.K. and Nomura, M.K.
TITLE Alkaline protease
JOURNAL Patent: EP 1347044-A 2 24-SEP-2003;
Kao Corporation (JP)
FEATURES
1..1305
Location/Qualifiers
1..1305
/organism="Bacillus sp. KSM-KP43"
/mol_type="unassigned DNA"
/db_xref="taxon:109322"
<1..1305
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAE85146.1"
/db_xref="GI:39922767"
/translation="NDVARGIVKADYVQSSVGLYQGOIVAVADTGLDTRGDSMHE
APRKITADYALGRINNDNGHGVHAGSVLNGSTNGKMGAPCANLVFOSIMDSGG
GLGGLPSLIQTLFSAQAKAIHTNSWGAAYVNGAYTDSNRVDDYVRKNDMTLFLFA
GNEGNGGTITFSAQAKAIHTNSWGAAYVNGAYTDSNRVDDYVRKNDMTLFLFA
DYNAQPTFLSASLAPDSFWANHSKYVMGTSNATPIVAGNVQLREHFVKR
GITPKPSLLKALIAADIGLYGPNQGWGRVTLDKSLNVAVNYESSSLSTOKAT
YSFTATAGKPLKISLVMSDAPASTTASVTLNLDLVLITAPNGTVYGVNDFTSPYNDN
WDGRNVENVFINAQSGTYTIEVQAYNVFVGPQFSLAIVN"
ORIGIN
Alignment Scores:
Pred. No.: 1,71e-133 Length: 1305
Score: 2247.00 Matches: 434
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-985-689A-1 (1-434) x AX839476 (1-1305)
Qy 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1 AATGATGTCGCGTGGGAATTTGCAAGCGGATGTGGCTCAGAGCAGCTACCGGTTGTAT 60
Qy 21 GlyGlnGlyGlnLeuValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGACAAGGACAGATCGTAGCGGTGCGGATACAGGCTTGATACAGTCCCAATGACAGT 120
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCAGTCATGAAGCCCTCCGCGGGAATTTACTGCAATTATATGCAATGGGACGCGAAT 180
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 181 AATGCCAATGATACGAATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 240
Qy 81 SerThrAsnGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 241 TCCACTAATAAAGAAATGGCGCTCAGCGGAATCTAGTCTTCCAAATCTATCATGTGATAGC 300
Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 301 GGTGGGGGACTTGGAGGACTACCTTGCATCTGCACAACTTATTCAGCCCAAGCATACAGT 360
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
Db 361 GCTGGTGCCAGAATTATACAAACTCCTGGGGAGCAGCAGTGAATGGGCTTACACACA 420

Qy 141 AspSerArgAsnValAspAspTyrValAlaArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 421 GATTCCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCTCTTTTGGTGGC 480
Qy 161 GlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 481 GGAATGAAGGACCGGAAACCGGACCATCATGTGACCGGACAGTCAAAATGAATA 540
Qy 181 ThrValGlyValAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 541 ACAGTCGGAGCTACGGAACCTCCGCCAAGCTTTGGTCTTATGCGGACAAATATCAAC 600
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 601 CATGTGGCAGTCTCTTCAGTGGACGCAAGGATGGCGGATCAAAACCGGATGTC 660
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 661 ATGGCACCGGAACTTCTATCTATCATGCAAGATCTTCTCTTGACCGGATCTCTCTTC 720
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 721 TGGCGCAACCATGACAGTAAATATGATGATGCGTGGTGGAACTCCATGCTACCGGATC 780
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 781 GTTCTGGAAACCTGGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCAATCACACA 840
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 841 AAGCTTCTCTATTAAGCGGCACTGATGCGGTCGAGCTGACATCGGCTTGGCTAC 900
Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 901 CCGAACGTTAACCAAGATGGGACGAGTGACATTTGGTAAATCTCTGACCTGACCTGCTAT 960
Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 961 GTGAACGAGTCCAGTCTCTATCCACGACCAAAAGCGACGTACTCGTTTACTGCTACT 1020
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1021 GCGGGCAAGCCTTTGAAATCTCCCTGGTATGTGTGATGCTGCTGCTGCTGCTGCTGCT 1080
Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1081 TCCGTAACTGTTGCAATGATCTGGACCTTGTCTATTACCGCTCCAAATGGCACACAGTAT 1140
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 1141 GTAGGAATGACTTACTTCGCCCATCAATGATACTGGGATGGCGCAATAACGATAGAA 1200
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1201 AATGATTTAATATGACCAACCAAGCGGACGCTATACATTCAGGTACAGGCTTATAAC 1260
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1261 GTACCGTGGACACACAGACCTTCTCGTTGGCAATTGTGAAT 1302

RESULT 2

AB051423

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1923 bp DNA linear BCT 10-MAY-2002

Bacillus sp. KP43 PROF gene for protease, complete cds.

AB051423

AB051423.2 GI:20521154

Bacillus sp. KSM-KP43

Bacillus sp. KSM-KP43

Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

1

Itoh, S. and Saeki, K.

TITLE new protease
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1923)
 AUTHORS Saeki, K.
 TITLE Direct Submission
 JOURNAL Submitted (21-NOV-2000) Katsuhisa Saeki, KAO CORPORATION;
 2606 AKABANE, ICHIKAIMACHI, HAGA, TOCHIGI 321-3486, Japan
 (E-mail: 3871850ka@saetnet.kao.co.jp, Tel: 81-285687471 (ex. 7471),
 Fax: 81-285687403)
 COMMENT On May 9, 2002 this sequence version replaced gi:14164344.
 FEATURES
 Location/Qualifiers
 1..1923
 /organism="Bacillus sp. KSM-KP43"
 /mol_type="genomic DNA"
 /strain="KP43"
 /db_xref="taxon:109322"
 1..1923
 /gene="PROF"
 1..1923
 /gene="PROF"
 /codon_start=1
 /transl_table=11
 /product="protease"
 /protein_id="BA55674.2"
 /db_xref="GI:20521155"
 /translation="MRKKKVFSLVLSAAAILSTVALSNPSAGGARNFOLDPKGIQTT
 TDAGFSKQOQTGAAPLVESENVKLPKGLQKLETVPANNKLHIITQFNGPILEETKQ
 QLEKTAGKLDIYIPDYAVIVEYEGDVKSATSTTEHVESVEPLYPIYRIDPOLFTKGAS
 ELVKVALDTPKQNKVQLRGIRIOIAQFALISNDVLYITAKPEVKWMDVARGIVKADV
 AQSSYGLYGGQIIVADVGLDTPGRNDSMHEAFRGKITLALVALGFTNNANDTNGHGT
 HVAGSVLNGSTNGKAPQANLVFQSIMSGGLGGLPNSQLTFSQATSAGARHIN
 SWGAANVAGTIDSRNVDDIVRKNDMTILFAAGNEFGNGTISAPGTAKNAITVGATE
 NLRPSGVADNINHAQFSRGGPTDKRIKPDVMAFGTIFLSARSLAPDSFWMNH
 DSKYAYMGTSMTPIVAGNVAQLREHFVKNRGTIPKPSLLKALIAAGADICLGPVN
 GNOGWGRVTLKSLNVAYNVNESLSSTOKATYSFTATAGKPLKISLVMSDAPASPTA
 SVTLVNDLIVITAPNGTQVVGNDFTSPYNDNDGNNVNFNAPQSGTITIEVQA
 YNPVPGPQTFSLAVN"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,76e-133 Length: 1923
 Score: 2247.00 Matches: 434
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 1 Gaps: 0
 US-09-985-689A-1 (1-434) x AB051423 (1-1923)
 Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrglyLeuTyr 20
 Db 619 AATGATGTTGGCGTGGAAATTGCAAGCGGATGTGGCTCAGAGCAGCTACGGGTTGTAT 678
 Qy 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGACAAGGACAGATCGTAGCGGTTGCCGATACAGGCGTTGATACAGTCCGCAATGACAGT 738
 Qy 41 SerMethHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGATGAAGCCTTCCCGGGGAAATTAACGCAATTATGCAATTGCGGACCGCAAT 798
 Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 799 AATGCCAATGATACGAATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 858
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 859 TCCACTAATAAGGAATGGCGCTCAGCGCAATCTAGTCTTCCAACTCTATCATGGATAG 918
 Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 919 GGTGGGGGACTTGGAGGACTTACCTTCGAATCTGCNAACCTTATTCAGCCAGCATACAGT 978

Qy 121 AlaGlyAlaAspGlyHisThrAsnSerTropGlyAlaAlaValAsnGlyAlaTyrThrThr 140
 Db 979 GCTGTGTCAGAAATTCATACAAATCTCTGGGAGCAGCAGTGAATGGGGCTTACACACA 1038
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1039 GATTCACAAAATGTGGATGATATGTGCGCAAAAATGATATGATGATGATGATGATGATG 1098
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1099 GGGAAATGAAGGACCGACCGCGGAACCATCATGTCACCGAGCAGCTTAAATAATGCAATA 1158
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1159 ACAGTCGGAGCTACGGAAAACCTCCGCCCAAGCTTTGGTCTTATGGGACAATATCAAC 1218
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 1219 CATGTGGCAGATGTTCTCTTCACGTGGACCGCAAGGATGACGAGTCAAAACCGGATGTC 1278
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 1279 ATGGCACCGGAAACGTTCTACTATCATCATCATCATCATCATCATCATCATCATCAT 1338
 Qy 241 TrpAlaAsnHisAspSerSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 Db 1339 TGGCGCAACCATGACAGTAAATATGTCATATGTCATGTCATGTCATGTCATGTCATGTC 1398
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 1399 GTTGTGGAAACGTTGGCACACTTCGTGAGCATTTGTGAAAACACAGAGGATCACACCA 1458
 Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1459 AAGCCTTCTCTATTAAAAGCGCACTGATTCGCGTGAGTGATGATGATGATGATGATGAT 1518
 Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1519 CCGAACCGTAAACCAAGGATGGGACGAGTGATGATGATGATGATGATGATGATGATGAT 1578
 Qy 321 ValAsnGluSerSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 1579 GTGAACGAGTCCAGTTCCTCTATCCACGACCCAAAAGACGAGTACTCGTTTACTGCTACT 1638
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrAla 360
 Db 1639 GCCGCAAGCCTTTGAAAATCTCCCTGGTATGGTCTGATGCGCTGCGAGCACCACTGCT 1698
 Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1699 TCCGTAAACGCTTGTCAATGATCTGGACCTTGTCAATTACCGCTCCCAATGGCACAGTAT 1758
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 Db 1759 GTAGGAAATGACTTTACTTCCCATACAAATGATACTGGATGGATGGCGCAATTAACGTAA 1818
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrThrIleGluValGlnAlaTyrAsn 420
 Db 1819 AATGATTTATTATATGACCAACAAAGCGGACGATATACAAATGAGGTACAGGCTTATAAC 1878
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1879 GTACCGGTTGGACACACAGACCTTCTCGTTGGCAATTTGTGAAT 1920
 RESULT 3
 AR368117
 LOCUS 1923 bp DNA linear PAT 12-SEP-2003
 DEFINITION Sequence 5 from patent US 6376227.
 ACCESSION AR368117
 VERSION AR368117.1 GI:34601778
 KEYWORDS Unknown.
 SOURCE Unknown.
 ORGANISM Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 1923)
 AUTHORS Takaiwa,M., Okuda,M., Saeki,K., Kubota,H., Hitomi,J., Kageyama,Y.,
 Shikata,S. and Nomura,M.

TITLE Alkaline protease

JOURNAL Patent: US 6376227-A 5 23-APR-2002;

FEATURES Location/Qualifiers

1..1923

source /organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
 Pred. No.: 2,76e-133 Length: 1923
 Score: 2247.00 Matches: 434
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-985-689A-1 (1-434) x AR368117 (1-1923)

Qy 1 AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 619 AATGATGTTGGCGTGGAAATTTGCAAGCGGATGTGGCTCAGAGCAGCTACCGGTTGTAT 678
 Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrClyArgAspSer 40
 Db 679 GGCAAGGACAGATCTAGCGGTTGCCGATACAGGCTTGATACAGTGCATACAGT 738
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGCATGAAGCCTTCGCGGGAATAATTACTGCATTATATGCAATGGGACGGAGAT 798
 Qy 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 799 AATGCCAATGATACGAATGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 858
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 859 TCCACTAATAAAGAAATGGCGCTCAGCGCAATCTAGTCTTCAATCTATCATGATAGC 918
 Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 919 GGTGGGGGACTTGGAGGACTACCTTCGAATCTGCAACCTTATTCAGCCAAACATACAGT 978
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAlaAsnGlyAlaTyrThr 140
 Db 979 GCTGGTGCCAGATTTCATAAACTCTGGGNGCAGCAGTGAATGGGCTTACACACA 1038
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1039 GATTCAGAAATGTGGAGTACTATGTGCGCAAAATGATATGACGATCCTTTTCGCTGCC 1098
 Qy 161 GlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1099 GGGAAATGAAGGACCGAAGCGGGAACATCATGTGCACAGGACAGCTTAAATGCAATA 1158
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1159 ACAGTCGGAGTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATCGGACAAATATCAAC 1218
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 1219 CATGTGGCAGTTCTTCTTACGTGGACCGCAAGAGATGGAGCGGATCAACCGGATGTC 1278
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
 Db 1279 ATGGACCGGAAAGCTTCTACTATCATCAGCAAGATCTTCTTTCGACCGGATTCCTTC 1338
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 1339 TGGGCGAACCATGACAGTAAATATGATCATGATGGTGGGAACGTCCTACCGATC 1398

Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 1399 GTTGCTGGAACCTGGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCATCACACCA 1458
 Qy 281 LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1459 AAGCCTTCTCTATTAAAGCGGCACTGATTCCGGTGACGTGACATCGGCCCTTGGCTAC 1518
 Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1519 CCGAACGTTAAACCAAGATGGGACGAGTGACATTGATAAATCCCTGAACGTTGCTAT 1578
 Qy 321 ValAsnGlnSerSerSerSerSerThrSerGlnLysAlaAlaThrTyrSerPheThrAlaThr 340
 Db 1579 GTGAACGAGTCCAGTCTCTCTATCCACAGCCAAAAGCGGACGTACTCTGTTACTGTACT 1638
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrAla 360
 Db 1639 GCCGGCAAGCCTTTGAAAATCTCCCTGGTATGCTCTGATGCCCTCGGAGCACACTGCT 1698
 Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1699 TCCGTAAACGTTGTCAATGATCTGGACCTTGTCAATTACCCCTCCAAATGSCACACAGTAT 1758
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnValGlu 400
 Db 1759 GTAGGAATGACCTTACTTCCGCATACATGATTAATGGATGCCGACATACAGTAGAA 1818
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1819 AATGATTTATTAATGACCAACAAAGCGGACGTATACATTTAGGTACAGGCTTAAAC 1878
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1879 GTACCGTTGACACAGACCTTCTCGTTGGCAATTGTGAAT 1920

RESULT 4

AB084155 1923 bp DNA linear BCT 02-SEP-2003
 DEFINITION Bacillus sp. KSM-9865 gene for protease, complete cds.
 ACCESSION AB084155

VERSION AB084155.1 GI:34392386

KEYWORDS

SOURCE Bacillus sp. KSM-9865

ORGANISM

Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE

1 Okuda,M., Saeki,K. and Kobayashi,T.

Bacillus sp. KSM-9865 protease gene

Published Only in Database (2003)

JOURNAL

2 (bases 1 to 1923)

REFERENCE

Okuda,M., Saeki,K. and Kobayashi,T.

Direct Submission

TITLE

Submitted (18-APR-2002) Mitsuyoshi Okuda, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail:okuda.mitsuyoshi@kao.co.jp,

Tel:81-285-68-7543, Fax:81-285-68-7547)

FEATURES

source

CDS

1..1923
 /organism="Bacillus sp. KSM-9865"
 /mol_type="genomic DNA"
 /strain="KSM-9865"
 /db_xref="taxon:192495"
 1..1923
 /codon_start=1
 /transl_table=11
 /product="protease"
 /protein_id="BAC82522.1"
 /db_xref="GI:34392387"
 /translation="MRKKKVFSLVSAAILSTVALSNPSAGSARNFOLDFKGIOTI
 TDAGFSKQGTGAAPLVESENKLPKGLQKLETFVANKLHIYQFNGPILERTKQ
 OLEKTGAKILDYIPDVAYIVEYEGDVKSATSTIEDVESVEPYLPFTRIDPQLFKGAS

ELKVALDITNOKNKEVOLREIOIAQFATNDVLYITAKPEYKWDVARGIVKADY
AQSYGLYGCCQVAIVADTGLDTRNDSSMHEAFPGKNTALTALGRNNANDINGHT
HVAGSVLNGSTNGKAPQANLVFQSIMDSGGGLGPLSNLTFLSQAYSAGARHGT
SWGAANGAYTTDSRNVDDYRKNMTILFAAGNEPGGT-SAPOTAKNAVITGATE
NLSPSGYADNTNHAQFSRSGPTKDRIKPDWAPAGTFFLSASSLAPDSFWANH
DSKYAYMGTSMTPTVAGNVASRLRHFVKNGRITPKPSLLKAALVAGAADIGLGYPN
GNQGWGRVTLDKSLNAVYNVESLSGTOKATVSTATAGKPLKLSLWSDAPASTTA
SVTLNDLDLIVITAPNGTQVGVNDFTSPYNNWDRNNVNFINAPQSGTIVIEVQA
YNYVPVGPQFSLAIVN"

ORIGIN

Alignment Scores:

Pred. No.: 5,72e-133 Length: 1923
Score: 2242.00 Matches: 433
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.78% Indels: 0
DB: 1 Gaps: 0

US-09-985-689A-1 (1-434) x AB084155 (1-1923)

Qy	1	AsnAspValAlaAaGgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr	20
Db	619	AATGATGTTGCAGTGGAAATGCAAGCGGATGGCGCAGACGCTACGGGTTGTAT	678
Qy	21	GlyGlnGlyGlnIleValAlaAaAspThrGlyLeuAspThrGlyArgAsnAspSer	40
Db	679	GGACAGGACAGATCGTAGCGGTTGCCGATACAGGCGTTGATACAGTGCATGACAGT	738
Qy	41	SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn	60
Db	739	TCGATGTCATGAAGCCTTCCGGGGGAAATACAGCATATATGATGATGGACGGAGAT	798
Qy	61	AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly	80
Db	799	AATGCCAATGATACGATGTCATGTCACGATGGCTCGCTCCGTATTAGGAACGGC	858
Qy	81	SerThrAsnIysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer	100
Db	859	TCCACTAATAAGAAATGGCGCTCTAGCGGAATCTAGTCTTCCAACTATCATGGATAGC	918
Qy	101	GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer	120
Db	919	GGTGGGGGACTTGGAGACATACCTTCGARTCGCAACCTTATTCAGCCAGCATACAGT	978
Qy	121	AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr	140
Db	979	GCTGTGCCAGAAATTCATCAAACTCTCGGGGAGCAGCAGTGAATGGGCTTACACAACA	1038
Qy	141	AspSerArgAsnValAspAspTyrValArgIysAsnAspMetThrIleLeuPheAlaAla	160
Db	1039	GATTCGAAATGEGGAGTACTATGCGCAAAATGATAGCATCCCTTTTCGGTGCC	1098
Qy	161	GlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaIysAsnAlaIle	180
Db	1099	GGGAATCAAGACCGAAGCGGGAACCATCAGTGCACCGGCACAGCTAAATAATGCAATA	1158
Qy	181	ThrValGlyAlaThrGluAsnLeuValGlyProSerPheGlySerTyrAlaAspAsnIleAsn	200
Db	1159	ACAGTCGGAGTACGGAAACCTCCGCCCAAGCTTCGGGTCTTATCGGGAATAATCAAC	1218
Qy	201	HisValAlaGlnPheSerSerArgGlyProThrIysAspGlyArgIleIysProAspVal	220
Db	1219	CATGTGGCACAGTTCTCTTCAGCTGGACCGCAAGAGATGACGGATCAACCGGATGTC	1278
Qy	221	MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe	240
Db	1279	ATGCAACGGGAGCGTTTCATACACTATCAGCAAGATCTTCTCTGACCGGATTCCTCTTC	1338
Qy	241	TrpAlaAsnHisAspSerIysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle	260
Db	1339	TGGCGGAACCATGACATGAATATGCAATACATGGTGGAAACGTCCTACGCCATC	1398

Qy	261	ValAlaGlyAsnValAlaGlnLeuArgGlyHisPheValIysAsnArgGlyIleThrPro	280
Db	1399	GTTGCTGGAACAGTGCGCACGCTTCGTGAGCATTTTGAAAAACAGAGGCATCACCA	1458
Qy	281	LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr	300
Db	1459	AAGCCTTCTCTATTAAAAAGCGGCACTGATTCGCGGTGCAGCAGACATCGGCTTGGCTAC	1518
Qy	301	ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspIysSerLeuAsnValAlaTyr	320
Db	1519	CCGAACGGTAACCAAGGATGGGACGAGTGATGGATAAATCCCTAAACGTTGCCCTAT	1578
Qy	321	ValAsnGluSerSerSerSerLeuSerThrSerGlnIysAlaThrTyrSerPheThrAlaThr	340
Db	1579	GTAACGAGTCCAGTTCCTCTATCCACGACCAAAAGCGAGCTACTCGTTTACTGCTACT	1638
Qy	341	AlaGlyLysProLeuLysIleSerIleuValTrpSeraspAlaProAlaSerThrAla	360
Db	1639	GCCGCAAGCCTTTGAAGATCTCCCTGGTATGGTCTGATGCCCTGGAGCACAACCTGCT	1698
Qy	361	SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlyTyr	380
Db	1699	TCCGTRACGCTTGTCATGACCTGGACCTTGCTGATTCATCCGCTCCAAATGGCACAATAT	1758
Qy	381	ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu	400
Db	1759	GTTGGAAATGACTTTTACTTCGCCATACAAATAATACTGGATGGGATGCCCAATAAGCTAGAA	1818
Qy	401	AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn	420
Db	1819	AATGATTTTATTATGCGCCCAAGCGGAGCATATACAAATGAGGTACAGGCTTATTAAC	1878
Qy	421	ValProValGlyProGlnThrPheSerLeuAlaIleValAsn	434
Db	1879	GTACCGGTGGACACACAGCCTTCTCGTTGGCAATTTGTGAAC	1920

RESULT 5

AR368118 1923 bp DNA linear PAT 12-SEP-2003
LOCUS Sequence 7 from patent US 6376227.
DEFINITION AR368118
ACCESSION AR368118
VERSION AR368118.1 GI:34601779
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1923)
Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y.,
Shikata, S. and Nomura, M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 7 23-APR-2002;
FEATURES Location/Qualifiers
source
1..1923
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 5,72e-133 Length: 1923
Score: 2242.00 Matches: 433
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 1
Query Match: 99.78% Indels: 0
DB: 6 Gaps: 0

US-09-985-689A-1 (1-434) x AR368118 (1-1923)

Qy	1	AsnAspValAlaAaGgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr	20
Db	619	AATGATGTTGCAGTGGAAATGTCAAAGCGGATGGCTCAGAGCAGCTACGGGTTGTAT	678
Qy	21	GlyGlnGlyGlnIleValAlaAaAspThrGlyLeuAspThrGlyArgAsnAspSer	40

679 GCACAAGGACAGATCGTAGCGTTCGCCGATACAGGCTTGATACAGGTCGCAATGACAGT 738
 Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGCATGAAGCCCTTCGCGGGAATAATCTCGCATATATATGCAATGGGACGAGCAAT 798
 Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 799 AATGCCAATGATACGAATGGTACGTCATGGTACGTCATGGTGGCTCGTATAGGAAACGGC 858
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 859 TCCACTAATAAAGGAATCGCGCTCAGCGGAATCTAGTCTTCCAAATCTATCATGGATAGC 918
 Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 919 GGTGGGGACTTGGAGGACTACCTCGAATCTGCAAACTTATTTCAGCCACGATACAGT 978
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
 Db 979 GTGGTGGCCAGAAATTCATCAAACTCTCGGGAGCAGCAGTGAATGGGGCTTACACAACA 1038
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1039 GATTCAGAAATGTGATGACTATGTGCGCAAAATGATATGACGATCTTTTCGCTGCC 1098
 Qy 161 GlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1099 GGGAAATGAAGGACCGAACCAGCCAGTACAGTGCAGCCAGGACAGCTAAAAATCAATA 1158
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1159 ACAGTCCGAGCTACGGAAACCTCGCCCAAGCTTTGGTCTTATGCGGACAAATATCAAC 1218
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 1219 CATGTGGCAGTCTCTCTTCAGCTGACCGACACAAAGAGTGGACGGATCAAAACCGGATGTC 1278
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 1279 ATGGCACCGGACGTTGATATATCATATCATGCAAGATCTTCCTTGCACCGGATTCCTCTTC 1338
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 Db 1339 TGGCGCAACCATGACAGTAAATATGATATGATGCTGGTGGAACTCCATGGCTACACCGATC 1398
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 1399 GTTGTGGAAACGTGGCAGCTTCGTGAGCATTTGTGAAACACAGAGGCATCACCA 1458
 Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1459 AAGCCCTTCCTAATAAAGCGGCACTGATTGCCGGTGCAGCTGACATGGCCCTTGGCTAC 1518
 Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1519 CCGAACGGTAACCAAGATGGGACGAGTGACATTGGATAAATCCCTGAACGTTCCTCAT 1578
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 1579 GTGAACGAGTCCAGTTCTCTATCCACACCGCAAAAGACGCTACTCGTTTACTGCTACT 1638
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 Db 1639 GCGCGCAAGCCCTTGAATAATCTCCCTGGTATGCTCTGATGCCCTCGCGACCAACTGCT 1698
 Qy 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1699 TCGGTACGGTGTGCAATGATCTGGACCTTGTCATTCACCTCCAAATGGACACAGTAT 1758
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 Db 1759 GTAGGAATGACCTTACTTCGCCCATACATGATGAATGATGCGGCGCGCAATACGTAGAA 1818

Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1819 AATGTTATTATTATATGACACCAACGCGGAGCTATACATTGAAGTACAGGCTTATAC 1878
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1879 GTACCGGTTGGACACAGAACTTCTCGTTGGCAATTTGTGAAT 1920

RESULT 6

AB046403 AB046403 1920 bp DNA linear BCT 10-MAY-2002
 LOCUS Bacillus sp. 9860 PROA gene for protease, complete cds.

DEFINITION AB046403

ACCESSION AB046403

VERSION AB046403.2 GI:20521152

KEYWORDS

SOURCE

ORGANISM

Bacillus sp. 9860

Bacillus sp. 9860

Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

gene

CDS

/organism="Bacillus sp. 9860"

/mol_type="genomic DNA"

/strain="9860"

/db_xref="taxon:133778"

1. 1920

/genes="PROA"

1. 1920

/genes="PROA"

/codon_start=1

/transl_table=11

/product="protease"

/protein_id="BAE21266.2"

/db_xref="GI:20521153"

/translation="MRKKVF.SVLAAAILSTVALNPNPSAGDARTFDLDFKGIOTTT

DVSGSKQRTGAAFLVSENVKLLKLETPANNKLIHVQNPGLILEETKQK

LEVTKAKILDYIPDVAIVVEGDIQSVRSIEHVESVEPYLPKYKIDPQLFTKGAST

LVKALADTKQNNKEVQLRIGIEIAQVANSVHVTITAKPEKVNVDVARGIVKADVA

QSYGLYGQGIIVAVADTGLDTRNDSSNHEAFRGKITAIYALGRNNANDINGRTH

VAGSVLGNATKNMAPQANLVFQSIIMDSGGGLGFLPSLQTLFQSFAGARITNS

WGAANVAGATYDTSRNDVTVRKNDMTILFAANGNERPNGGTISAPGTAKNAITVGATN

LRPSFGSYADNINHVAQFSRGTQDKRIKPDVMAPTGYILSARSLAPDSFFWANHD

SKYAVMGSTSMATPIVAGNVAOLREHFVNKRGITPKPSLLKALTAGAADVGLYPNG

NCGMGRVTLQSLNVAVNVNESSALSTSQKATVFTATAGKPKLSILVNSDAPSTTAS

VTLVNDLDELVITAPNGTRYGNDPFSAPFNNDGNNVENVFINSQSQTYYIEVQAY

NVPVPGQNFSLAVN"

ORIGIN

Alignment Scores:

Pred. No.: 3,09e-129 Length: 1920

Score: 2183.00 Matches: 418

Percent Similarity: 99.31% Conservative: 13

Best Local Similarity: 96.31% Mismatches: 3

Query Match: 97.15% Indels: 0

DB: 1 Gaps: 0

US-09-985-689A-1 (1-434) x AB046403 (1-1920)

QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 616 AATGATGTGGCCAGAGGTTATGTCAAAGCGGATGTGCCACAGCAGCTACGGTTTGTAT 675
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGACAAAGGCCAGATTCGCGAGTTGCCGATATCTGGATTGGATACAGGAAGAACACACAGT 735
QY 41 SerMetHisGluAlaPheArgGlyIysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATCATGAGCCCTTCGCGGTAAATAACAGCACTATATGCACTGGTGGCGAGAT 795
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 796 AATGCCAATGATACGAAACCGTATGTCACCCATGTGGCAGGTTCGGATATAGGAATGGC 855
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 856 GCAACGAAATAAAGGATGACCTCAAGCGATCTGGTTTTCATCCATCATGATAGC 915
QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 916 AGTGGTGGGCTTGGAGGCTTGCCTTCCAAATCTGCAACCTTATTCAGCCAGCATTCAGT 975
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
Db 976 GCAGGTGCCAGATTCATCAAACTCTCGGGGGCAGCGGTGATGGGGCTTACAGACA 1035
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1036 GATTCAGAAATGTGGATGACTATGTAAGGAAAAATGATACGATTCATTCCTTCGGGT 1095
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1096 GGGAAATGAAGCGCGAAGCGCGTATCATCAGTGCACCTGGTACGGCTAAACGCCATA 1155
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1156 ACAGTGGCGCAACCGAAACCTCGCTCAAGCTTCGGTTCTTATGCAGATAATATTAAAC 1215
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1216 CACGTTGCACAGTTCTCTTCCCGTCCCGACAAAAGATGGCGAATCAAGCCTCATGTC 1275
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 1276 ATGGCGCCAGGACATACATTTTATCAGCAAGATCTTCTTGCACCCGATTCCTCCTTC 1335
QY 241 TrpAlaAsnHisAspSerIysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
Db 1336 TGGCGCAATCATGACAGCAAAATATGCTATATGGGTGGAACGTCCATGGCAACACCGAT 1395
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyIleThrPro 280
Db 1396 GTTGGCGGGAATTTGCACAGCTCCGTCAGCATTTTGTGAAATAACAGAGATCACTCCT 1455
QY 281 LysProSerLeuLeuIysAlaLeuIleAlaGlyAlaAspIleGlyLeuGlyTyr 300
Db 1456 AAGCCTTCCCTATTGAAAGCAGCTTTGANTGCAAGTGTCTGCTATGTTGGATTGGGTAT 1515
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1516 CCGAAGCAACCAAGATGGGCGCAGTGACCTCGATAAATCGTTGAACTGTGGCTAT 1575
QY 321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 1576 GTGAACGAATCCAGTGCCCTTCACTAGCCAAAGAGACATATACCTTTACTGCAAG 1635
QY 341 AlaGlyLysProLeuIysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1636 GCGGGCAAGCCATTTGAAATCTCCCTGGTATGTGTCGGATGCCCTTCAAGCACTACTGCT 1695

QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1696 TCTGTAAACCTCGTCAATGATTTGGATTTCGTCATTACAGCACCAACGGAACAGATAT 1755
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 1756 GTCCGGGAATGACTTCTCGACCCANTTGCACAACTGGGATGGCCCAATACGTAGAA 1815
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1816 AATGTAATTTAATTCGCCCAAGTGGAAACATATACCATTCAGTGTCAAGCATATAAT 1875
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1876 GTCCGGTTGGACCAAAAACCTTCCTGTTGGCAATTTGTGAAC 1917
RESULT 7
AR368116 AR368116 1920 bp DNA linear PAT 12-SEP-2003
LOCUS AR368116 Sequence 3 from patent US 6376227.
DEFINITION AR368116
ACCESSION AR368116
VERSION AR368116.1 GI:34601777
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1920)
AUTHORS Takaiwa,M., Okuda,M., Saeki,K., Kubota,H., Hitomi,J., Kageyama,Y.,
Shikata,S. and Nomura,M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 3 23-APR-2002;
FEATURES Location/Qualifiers
source 1..1920
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores: 3.09e-129 Length: 1920
Pred. No.: 2183.00 Matches: 418
Score: 99.31% Conservative: 13
Percent Similarity: 96.31% Mismatches: 3
Best Local Similarity: 97.15% Indels: 0
Query Match: 6 Gaps: 0
DB:
US-09-985-689A-1 (1-434) x AR368116 (1-1920)
QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 616 AATGATGTGGCCAGAGGTTATGTCAAAGCGGATGTGCCACAGCAGCTACGGTTTGTAT 675
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGACAAAGGCCAGATTCGCGAGTTGCCGATATCTGGATTGGATACAGGAAGAACACAGT 735
QY 41 SerMetHisGluAlaPheArgGlyIysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATCATGAGCCCTTCGCGGTAAATAACAGCACTATATGCACTGGTGGCGAGAT 795
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 796 AATGCCAATGATACGAAACCGTATGTCACCCATGTGGCAGGTTCGGATATAGGAATGGC 855
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 856 GCAACGAAATAAAGGATGACCTCAAGCGATCTGGTTTTCATCCATCATGATAGC 915
QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 916 AGTGGTGGGCTTGGAGGCTTGCCTTCCAACTGCAACCTTATTCAGCCAGCATTCAGT 975
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140

Db 976 GCAGGTGCAGAAATTCATACAAATCTCTGGGGGAGCGGTGAATGGGGCTACACGACA 1035

Qy 141 AspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeuPheAlaala 160

Db 1036 GATTCAGAAATGTGGATGACTATATAGGAAATATGATAGCATTTCTTTTCGGCGCT 1095

Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaile 180

Db 1096 GGGATGAAAGCCGAAACGGGGTACCATCATGTCACCTGTCAGCGCTAAAGACGCATA 1155

Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleasn 200

Db 1156 ACAGTCGGCGCAACCGAAACCTGGTCCAGCTTCCTGCTTCTATGACAGATAATTAAC 1215

Qy 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220

Db 1216 CACGTTGCACAGTCTCTTCCGTGGCGCGACAAAGATGGCGAATCAACSCCTGATGTC 1275

Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240

Db 1276 ATGGCGCCAGGACATACATTTATCAGCAAGATCTTCTTGCACCGGATTCCTCTTC 1335

Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProile 260

Db 1336 TGGCGAATCATGACAGCAATATGCTATATGGTGGACGTCCATGGCAACCGATT 1395

Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280

Db 1396 GTTGGCGGAATGTTGCACAGCTCGTGAGCATTTTGTGAAATAATAGAGCAATCACTCT 1455

Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300

Db 1456 AAGCCTTCCTATTGAAGACCTTTGATGACAGTGTCTGTATGTTGGATTGGGTTAT 1515

Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320

Db 1516 CCGAACCGAAACCAAGGATGGGGCGAGTGACCTGGATAAATCGTTGAACGTTGCCTAT 1575

Qy 321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340

Db 1576 GTGACGAATCCAGTGGCCCTATCAATAGCCAAAGGACATATACCTTTACTGCAACG 1635

Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360

Db 1636 CGGGCAAGCCATTCGAAATCTCCCTGGTATGTCGGATGCCCTGCAAGCACTACTGCT 1695

Qy 361 SerValThrLeuValAsnAspLeuAlaIleThrAlaProAsnGlyThrGlnTyr 380

Db 1696 TCTGTAAACCTCGTCAATGATTGGATTGGTCAATACAGCACCACCAACGGAAGATAT 1755

Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400

Db 1756 GTCCGGATGACTTCTCAGCACCAATTGACATAACTGGGATGGCGCGCAATAACGTAGAA 1815

Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420

Db 1816 AATGTATTATTATTCGCCCCAAGTGGAAACATATACCATTCAGGTGCAAGCATATAAT 1875

Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434

Db 1876 GTGCCGGTGGACCAACAACTTCGTTGGCAATTGTGAC 1917

RESULT 8

AB046406
LOCUS
DEFINITION Bacillus sp. NV1 PROE gene for protease, partial cds.
ACCESSION AB046406
VERSION AB046406.1 GI:12381944
KEYWORDS
SOURCE Bacillus sp. NV1
ORGANISM Bacillus sp. NV1
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE

AUTHORS
Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and Horikoshi, K.

TITLE

Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic *Bacillus* spp.: enzymatic properties, sequences, and evolutionary relationships

JOURNAL

Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

MEDLINE

20568675

PUBMED

11118284

REFERENCE

2 (bases 1 to 1302)

AUTHORS

Saeki, K.

TITLE

Direct Submission

JOURNAL

Submitted (20-JUL-2000) Katehisa Saeki, Kao corporation, Biological Science Laboratory, Ichikaimachi Akabane 2806, Hoga, Tochigi 321-3497, Japan (E-mail: 387185@kasanet.kao.co.jp, Tel: 81-285-68-7400, Fax: 81-285-68-7403)

FEATURES

source

1..1302

/organism="Bacillus sp. NV1"

/mol_type="genomic DNA"

/strain="NV1"

/db_xref="taxon:133781"

1..1302

/gene="PROE"

<1..>1302

/gene="PROE"

/codon_start=1

/transl_table=11

/product="protease"

/protein_id="BAB21269.1"

/db_xref="GI:12381945"

/translation="NDVARGIVKDAQSSYGLYGGQVAVADTGLDTRNDSSMHE
AFRGKTAIYALGRNTNNDPNHGHTHVAGSVLNGTSNKGMAPQANLVFQSVDSNG
GLGGPSNVSTLPSGAYSAGAPIHTNSGAPVNGAYTTDSRVDDYVRKNDMAVLFAA
GHEPNGGTTISAPKAKNAITVGATENLRPSGSYADNINHVAQSSRGPTKDGILKP
DYNAPGTFILSARSLAPDSFWANHSKIATMGTSMTATPIVAGNVAQREHFIGNR
GITPELSLKAALAGATDILGYPGSGQNGRVTLDKSLNVAFNVETSSSTNQAT
YFTAQSGKPLKISLWSDAPATASVTLVNDLVLITAPNGTKYGVNDFTPYDNN
WDGRNVENVFINAQSGTYTVEVQAYNVQPGQAFSLAIVN"

ORIGIN

Alignment Scores:

Pred. No.: 6,49e-127 Length: 1302
Score: 2143.00 Matches: 406
Percent Similarity: 97.93% Conservative: 19
Best Local Similarity: 93.55% Mismatches: 9
Query Match: 95.37% Indels: 0
DB: 1 Gaps: 0

US-09-985-689A-1 (1-434) x AB046406 (1-1302)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20

Db 1 AACGATGTTGCCAGAGCGCATTTGTAAAGCCGATGTTCCAGAGCAGCTATGGTTATAT 60

Qy 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrClyArgAsnAspSer 40

Db 61 GGGCAAGGCGCAAGTGTTCAGTAGCGGATACCGGACTGGATACAGGGCGTATGACAGC 120

Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60

Db 121 TCATGTCATGACGCTTCCGAGGAAGATTACCGGATATATGCCCTTGGAAAGACAAAC 180

Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80

Db 181 AACGCCAATGATCCAAACGACACGCGGACGCGCATGTTGCCGGATCTGTTTATGCAACCGT 240

Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100

Db 241 ACTTCGAATAAGGATGGCTCCACAGACTACTTAGTTTCCATCTCTGTTATGACAGC 300

Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120

301 AATGGCGGACTTGGCGGACTGCTTCCAAATGTAAGTACATATTATTCAGCCAGGCATATAGT 360
121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValAlaValAsnGlyAlaTyrThr 140
361 GCCGGTGCAGAAATCCATAGAACTCATGGGAGCGCCTGTAAACGGAGCGTACATCT 420
141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAla 160
421 GATTCCAGAAACGTAGACGATTATGTTGGTAAATGATATGATGGCGTTCTTTTTCAGCG 480
161 GlyAsnGluClyProAsnGlyThrIleSerAlaProGlyThrAlaLysAlaIle 180
481 GGTAAACGAGGCGGAATGCGGCACAAATCAGTCTCCTGGTACCGGAGAAATGCTATC 540
181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
541 ACAGTAGGGCAACAGAAAACCTGGGCCAAGCTTTGGATCTTATGCTGACAACTCAAT 600
201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
601 CATGTAGCACAGTTTCTCCCGGACCTTACAAAGGATGGAGTATCAACCGGACGTA 660
221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
661 ATGGCACCAGCAATTTATTTATCGCAAGATCTTCTTGGCTCGGACTCTCTCATTC 720
241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyClyThrSerMetAlaThrProIle 260
721 TGGCAAAACCATGACAGCAATATGCTTATATGGTGGAACTCCATGGCGACACCGATT 780
261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
781 GTAGTGGTAACGTTGGCAGGTACGTGAACATTTATCMAAAACAGAGAAATCATCTCT 840
281 LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
841 AAACCATCTTGTGAAAGCAGCTCTATTGCCGAGCAACTGATATCGGTCTTGGCTAT 900
301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
901 CCGAGTGGAAACCCAGGATGGGAAGAGTAACATTTGACAAAGTCACTTAATGAGCTTTC 960
321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
961 GTAAATGAAACAAGCTCTTTATCTACTAACCAAAAGGCTACGTATTCTATTCTAGCACA 1020
341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrAla 360
1021 TCAGGCAAAACCTTTGAAGATTTCAATGGTTTGTGTGATCCACCGGCAAGTACTCCGCA 1080
361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
1081 TCGGTTACATTTGGTGAATGATCTGGATCTGGTGTATACAGTCCAAATGGAAAGATAT 1140
381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
1141 GTTGGAAACGACTTACTGCTCCCTATGATAATAACTGGGATGGAGTACCAATGTAGAG 1200
401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
1201 AACGTGTTTATCAATGCTCCGCAAGCGGAACGTATACAGTTGAGGTACAGGCTTACAT 1260
421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
1261 GTTCCCAAGGGCGGCGGCTTTCTTTTGGCTATTGTGAAC 1302

RESULT 9
AR069954
LOCUS AR069954 3003 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 41 from patent US 5891701.
ACCESSION AR069954
VERSION AR069954.1 GI:7220842
KEYWORDS

SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3003)
AUTHORS Sloma,A. and Christianson,L.
TITLE Nucleic acids encoding a polypeptide having protease activity
JOURNAL Patent: US 5891701-A 41 06-APR-1999;
FEATURES
Location/Qualifiers
source 1..3003
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Alignment Scores:
Pred. No.: 2,34e-125 Length: 3003
Score: 2125.50 Matches: 406
Percent Similarity: 97.93% Conservative: 19
Best Local Similarity: 93.55% Mismatches: 8
Query Match: 94.59% Indels: 1
DB: 6 Gaps: 1
US-09-985-689A-1 (1-434) x AR069954 (1-3003)
QY 1 AsnAspValAlaAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 1470 AATGACGTGGCGCTGGCATTTGTGAAGCAGACGCTGCACAAATAACTTTGGCTTATAT 1529
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 1530 GGACAAGGACAGATTGTAGCAGTTGCTCATCTGGCTTGATACAGGAAGAATGACAGT 1589
QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 1590 TCGATGCAATGAAGCATTCGCGGTAAAGATTACCCACTATATGCTGCGCGAAGCAAGAT 1649
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 1650 AACGCCAATGATCCAAATGGACATGGAAACCCATGCTGCTGGATCTGTGTAGGAAT--- 1706
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 1707 GCTCAAAATAAAGGATGGCACCCAGCAAGCCAAATCTAGTCTTTCAATCTATTATGGATAGT 1766
QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
DB 1767 GGTGGAGGCTGGGAGGACTACCTGCTAATCTACAACTATTTCAGTCAAGCATATAGT 1826
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
DB 1827 GCTGGAGCGAGAAATTCATACGAATTCATGGGGGCTCCAGTAAACGGTGCCTATACGACA 1886
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 1887 GACTCTCGAATTTGTGATTTATGTGAGAAAATAATGATATGACGATCTTTTTGGGCGC 1946
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAlaIle 180
DB 1947 GGAAATGAGGACCAAGGAGTAGCGGTACAAATCAGTCCAGGAAACAGCAAAATAATGCGATT 2006
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
DB 2007 ACAGTTGGGGCAACCGAAAACCTTACGTCCAAAGCTTCGATCTTATGGGATATATTAAC 2066
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
DB 2067 CATGTTGCTCAATCTCTTCACGAGGCTCTACTAGATGAGCGATTAATTAAGCCGAGCTC 2126
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
DB 2127 ATGGCACCAGGTAGCATATATTTCTCTGCTAGATCATCATTAGCTCAGATTTCTTCATTC 2186
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260

Db 2187 TGGGCAACCAATGATAGTAATAATGCTACATGGGTGGTACTTCTATGCTACTCCAAATT 2246

Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyIleThrPro 280

Db 2247 GTAGCAGGTAATGTTGCCACAAATTAAGGAGCAATTTGTGAAAATAGAGGGTAATCTCT 2306

Qy 281 LysProSerLeuLeuLysAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300

Db 2307 AAGCCCTTCCTTTAAAGCTGCTTTAAATGTCAGGTGCTGCGGATGTTGGAGCTTGGCTTT 2366

Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320

Db 2367 CMAATGGTAACCAAGATGGGAAGAGTAACGTTAGATAAATCCCTAAATGTCGCAATT 2426

Qy 321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340

Db 2427 GTGAATGAACAGAGCCCTTTATCAACAGTCAAAAAGCAACATATTCGTTACGGCTCAA 2486

Qy 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrAla 360

Db 2487 GCTGGTAACCCCTTAATAATATCACTTGTGTCAGATGCACCAAGTAGACACGCGCA 2546

Qy 361 SerValThrLeuValAsnAspLeuLeuValIleThrAlaProAsnGlyThrGlnTyr 380

Db 2547 TCACTAACTTTAGTAATGATTTAGACTTAGTAATCACAGCAACCAATGGAATCAATAC 2606

Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnValGlu 400

Db 2607 GTCGGAATGACCTTTACAGACCCGATGATACAAATTTGGATGGCAGAAACAACTGGAA 2666

Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420

Db 2667 AATGCTTTATCAATCTCTCAAGCGGAACGTATACAGTCAAGTCAAGTCAAGCTTACAAT 2726

Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434

Db 2727 GTACCAAGTAGTCCGCAACCTTTCTTGGCATTTGATACAT 2768

A25

RESULT 10

LOCUS BD062155 3003 bp DNA linear PAT 27-AUG-2002

DEFINITION Nucleic acids encoding a polypeptide having protease activity.

ACCESSION BD062155

VERSION BD062155.1 GI:22607760

KEYWORDS JP 2001514529-A/39.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1. (bases 1 to 3003)

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE Sloma, A. and Christianson, L.

JOURNAL Nucleic acids encoding a polypeptide having protease activity

NOVO NORDISK BIOTECH INC

COMMENT PN JP 2001514529-A/39

PD 11-SEP-2001

PF 09-JUN-1998 JP 1999503145

PR 12-JUN-1997 US 08/873479

PI ALAN SLOMA, LYNNE CHRISTIANSON

PC C12N15/57, C12N15/75, C12N9/54, C12K14/00

CC Strandedness: Single;

CC Topology: Linear;

FH Key Location/Qualifiers.

FEATURES

source 1. .3003

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 2,34e-125 Length: 3003

Score: 2125.50 Matches: 406

Percent Similarity: 97.93% Conservative: 19

Best Local Similarity: 93.55% Mismatches: 8

Query Match: 94.59% Indels: 1

DB: 6 Gaps: 1

US-09-985-689A-1 (1-434) x BD062155 (1-3003)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20

Db 1470 AATGACGTGGCCGTGGCAATTGTCAAGACAGACGTGCACAAATAACTTTGGCTTATAT 1529

Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40

Db 1530 GGACACAGACAGATTTGAGACGTTGCTGATACGTGGCTTGATACAGGAAGAAATGACAGT 1589

Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60

Db 1590 TCGATGATCAAGCAATTCGCGGTAAAGATTACGCACTATATGCACCTGGGACAGCAAT 1649

Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80

Db 1650 AACCCCAATGATCAAAATGGACATGGAAACCAATGTTGCTGGATCTGTGTAGGAAT--- 1706

Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100

Db 1707 GCTACAAATAAGAGGATGGCACCGCAAGCCAACTCTAGTCTTTCAATCTATTATGGATAGT 1766

Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120

Db 1767 GGTGGAGGGCTGGAGGACTACCTGCTAATCTACAACATTTATTCAGTCAAGCATATAGT 1826

Qy 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyValAlaValAlaAsnGlyAlaTyrThr 140

Db 1827 GCTGACGAGAGATTTCATCAATTCATGGGGGCTCCAGTAACAGTGCCTATACGACA 1886

Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160

Db 1887 GACTCTCGAAATGTTGATGATTATGTAGAAAAAATGATATGACGATTTCTTTTGGGGCC 1946

Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180

Db 1947 GGAATGAGGACCAAGTACGGGTACATCAGTGCACCAAGACAGCAAAAAATGCGATT 2006

Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200

Db 2007 ACAGTTGGGCAACCCGAAAAACCTACGTCCAAGCTTCGGATCTTTATCGGATAATATAAC 2066

Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220

Db 2067 CATGTTCTCAATTCCTTCCAGAGTCTCTAGAGTGGACGTATTAGCCGAGCGTC 2126

Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240

Db 2127 ATGGCACCAAGTACGTATATTTCTCTGTAGATCATCATTAGCTCCAGATTCCTCATTC 2186

Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260

Db 2187 TGGCAAAACCATGATAGTAATAATGCTCATACGCGGTACTTCTATGGCTACTCCAAAT 2246

Qy 261 ValAlaGlyAsnValAlaGlnLeuAspGluHisPheValLysAsnArgGlyIleThrPro 280

Db 2247 GTACAGGTAATGTTGCACAAATAAGGAGCAATTTGTGAAAAAATAGAGGGTAATCCT 2306

Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAspIleGlyLeuGlyTyr 300

Db 2307 AAGCCTTCCCTTTTAAAGCTGCTTTAATGTCAGGTGCTCGGATGTGGACTTGGCTTT 2366

Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320

Db 2367 CCAATGGTAACCAAGGATGGGAGAGTAACGTTAGATAAATCCCTAAATGTCGCAATTT 2426

Qy 321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340

Db 2427 GTGAATGAACGAGCCCTTTATCAACAAGTCAAAAGCAACATATTCGTTACGGCTCAA 2486

Qy 341 AlacGlySerProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360

Db 2487 GCTGGTAACCCCTTAAATAATACACTTGTGTGGTCAGATGACACGAGTAGCAGCGCA 2546

Qy 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380

Db 2547 TCACAACTTTAGTGAATGATTTAGACTTAGTAACTCACTGCACCAATGAACATAATAC 2606

Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400

Db 2607 GTCGGAATGACTTTACAGCACCGCTATGATAACAAATGGGATGGCAGAAACAACGCGAA 2666

Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420

Db 2667 AATGTGTTTATCAATGCTCCTCAAGCGGAAGCTATACAGTCGAAGTGCAGGCTTACAT 2726

Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434

Db 2727 GTACCAGTAAGTCGCGCAACCTTTCTTTAGCGATTGTACAT 2768

RESULT 11

AB046405

LOCUS AB046405 1299 bp DNA linear BCT 23-JAN-2001

DEFINITION *Bacillus* sp. SD521 PROD gene for protease, partial cds.

ACCESSION AB046405

VERSION AB046405.1 GI:12381942

KEYWORDS

SOURCE *Bacillus* sp. SD521

ORGANISM *Bacillus* sp. SD521

REFERENCE 1 (bases 1 to 1299)

AUTHORS Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and Horikoshi,K.

TITLE Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic *Bacillus* spp.: enzymatic properties, sequences, and evolutionary relationships

JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

MEDLINE 20568675

PUBMED 11118284

REFERENCE 2 (bases 1 to 1299)

AUTHORS Saeki,K.

TITLE Direct Submission

JOURNAL Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation, Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga, Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp, Tel:81-285-68-7400, Fax:81-285-68-7403)

FEATURES

source

1..1299

/organism="Bacillus sp. SD521"

/mol_type="genomic DNA"

/strain="SD521"

/db_xref="taxon:133780"

1..1299

/gene="PROD"

<1..>1299

/gene="PROD"

/codon_start=1

/transl_table=11

/product="protease"

/protein_id="BAB21268.1"

/db_xref="GI:12381943"

/translation="NDVARGIVKADVQNNYGLYGQGVVAVADTGTIDGRNDSMHE AFKRIKIALVAGRTNNDPNHGHTVAGSVLGNALNKGMAQOANLVFOSIMDSGG LGGPISNLNTLFSQAWNAGARLHTNSWGAIPNGAYTANSFQVDEYVRNNDMTVLPAAG NEGNSGTSIAPGTAIXALIVGATENYRPFSGSLADPNPHIAQFSRGATRDRIKRP VTAPGTFILSGSLADPDSFWANYSKIYMGTSWATFVAGNVAQUREHFIRKRG ITPKPSIKKALTAGTAVGLVPSGQDQGWRLDKSLNVAYVNEATLALATQKATY SFOAQAGKPLISLVTDAPGSTASTYLLVNDLDELVITAPNGQKTYGNDPSYPIDNW DGRNNVENVFINAPQSQGTYTEVQAVNVPSPGQRFSLAIHV"

gene

CDS

ORIGIN

Alignment Scores:

Prod. No.: 9, 02e-118 Length: 1299

Score: 1998.50 Matches: 381

Percent Similarity: 94.24% Conservatives: 28

Best Local Similarity: 87.79% Mismatches: 24

Query Match: 88.94% Indels: 1

DB: 1 Gaps: 1

US-09-985-689A-1 (1-434) x AB046405 (1-1299)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20

Db 1 AATGATGTAGCAAGAGGAATAGTAAAGACGACGCTGTGCACAAAACAATACGGACTATAT 60

Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40

Db 61 GGACAAGGTCAAGTAGTTGCAGTAGCGGATACGGGTTTAGATACAGGTCGTAAACGATAGT 120

Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60

Db 121 TCTATGTCATGAAGCATTCCTGGGAAAATTAACAGCTCTTTACGGCTTAGGAAGAACTAAC 180

Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80

Db 181 AATGCAATGATCCGAATGGGCATGTCAGCATGTAGCAGGTTCTGTACTTGGTAAT--- 237

Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleValMetAspSer 100

Db 238 GCTTTAAATAAAGGAATGGCTCCGCAAGCTAACTTAGTCTTCCCAATCTATTATGATGATAGC 297

Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120

Db 298 ACGGAGGATTAGGTGGATTTACCATCGAATTTGAATAGCTATTATTAGTCAGAGCTGGAAAT 357

Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140

Db 358 GCTGGGGCTAGAAATTCATACTAACTCTTGGGGTGTCCAGTAAATAGGAGCGGTACACTGCT 417

Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160

Db 418 AACTCGACAAAGTGGATGAGTATGTTCCAAATAATGATATGACGCTACTTTTTCGACGA 477

Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180

Db 478 GGTAAATGAAGTCCTCTAATTCAGAAACAATTAGTGTCTCCAGCACAGCGAAAAATGCCATT 537

Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200

Db 538 ACGGTCCGCGCAACGGAATACTATCCCGCGAGCTTCGGTTTCATTAGCAGATAACCCAAAT 597

Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220

Db 598 CATATCGCACAAATTTTCATCAAGAGGAGCTACGAGAGATGGACGAATTAACCGACGTA 657

Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240

Db 658 ACAGCTCTCTGGAAACATTTATTTTATCAGCACGCTTCTTCTTAGCCCCAGACTCTTCGTTT 717

Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260

Db 718 TGGGCGAATTAACAGTAAGTATCGGTACATGGCGGTACTCTATGCGGCACACCTATA 777

Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280

Db 778 GTTGGGGGAATGTCCGCAATTTAGCGGAGCATTTTATAAAAAATAGAGGAATTAACACT 837

Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300

Db 838 AAACCTTCTTAAATAAAGCTGCACTTATCGCTGGGGCTACTGATGTTGGTTAGGATAT 897

Qy 301 ProAsnGlyAsnGlnGlyTyrTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320

Db 898 CCAAGTGTGTGACCAAGGCTGGGGCGTGTACTTACTAGTATAAATCGTTAAATGTAGCGTAT 957

QY 301 ProAsnGlyAenGlnGlyTTPGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 898 CCAAGTGGTACCAAGGCTGGGGCGGTACTTTAGATAAATCGTTAAATGTAGCGTAT 957
 QY 321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 958 GTCAATGAAGCACTGCATTAAACACAGGACAAAGAACAGCTATTCTGCCAACGCCAA 1017
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTTPSerAspAlaProAlaSerThrAla 360
 Db 1018 GCGGGTAAACCATTAATAATCTGCTAGTAGTGACAGATGCACCTGGAAGTACACAGCA 1077
 QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1078 TCTTATACACAGTAAATGATTTAGATCTAGTATTACTGCTCGAATGGACAAAATAT 1137
 QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTTPAspGlyArgAsnValGlu 400
 Db 1138 GTAGTAATGATTTAGTTATCTTATGATAATTAATTTGGATGGTCCACAAATGTTGAG 1197
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1198 AACGTATTATTAAACGCTCCGCAATCTCGAACGTATACAAATTCAGGTTCGAAGCGTATAAC 1257
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1258 GTTCATCTGCACACACAGCGTTCTCCTACAGTATCGTATAT 1299

RESULT 13
 AB046404
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 JOURNAL
 TITLE
 JOURNAL
 TITLE
 JOURNAL
 FEATURES
 source
 gene
 CDS

AB046404 1299 bp DNA linear BCT 23-JAN-2001
 Bacillus sp. Y PROC gene for protease, partial cds.
 AB046404.1 GI:12381940
 Bacillus sp. Y
 Bacillus sp. Y
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 1 (sites)
 Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H. and
 Horiuchi K.
 Novel oxidatively stable subtilisin-like serine proteases from
 alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
 evolutionary relationships
 Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
 20568675
 11118284
 2 (bases 1 to 1299)
 Saeki K.
 Direct Submission
 Submitted (20-JUL-2000) Katsuhisa Saeki, Kao Corporation,
 Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
 Tochigi 321-3497, Japan (E-mail:387185@kasta.net.kao.co.jp,
 Tel:81-285-68-7400 Fax:81-285-68-7403)
 Location/Qualifiers
 1..1299
 /organism="Bacillus sp. Y"
 /mol_type="genomic DNA"
 /strain="Y"
 /db_xref="taxon:133779"
 1..1299
 /gene="PROC"
 <1..>1299
 /gene="PROC"
 /codon_start=1
 /evidence=experimental
 /product="protease"
 /protein_id="BAB21267.1"
 /db_xref="GI:12381941"
 /translation="NDVARGIVKADVAQNNYGLGQGVVAVADTGLDTRNDSSMHE
 AFRKITALVALGNTNADPNHGHTVAGSVLGNALNKGMAPCANLVFOSIMDSGG
 LGGLPSNLTLFSAQWAGARIHTNSWGAIPVNGAYTANSRQVDEYVRNNDMTVLFAAG

ORIGIN

Alignment Scores:

Pred. No.: 4,486-117 Length: 1299
 Score: 1987.50 Matches: 379
 Percent Similarity: 94.01% Conservative: 29
 Best Local Similarity: 87.33% Mismatches: 25
 Query Match: 88.43% Indels: 1
 DB: 1 Gaps: 1

US-09-985-689A-1 (1-434) x AB046404 (1-1299)

QY 1 AsnAspValAlaAArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 1 AATGATGTAGCAAGAGGATAGTAAAGCTGATGTTCACAAACAATACGGATTATAT 60
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 61 GGACAAGGTCAGTAGTGTTCAGTAGCGACACAGCTTAGATACAGGTCGTAAACGATAGT 120
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGTCATGAAGCATTCGCGGGAATAATCAGCTCTTTACGCGTTAGGAAGAATAAT 180
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 181 AATGCGAGTGATCCGATGCGCATGCGACACATGTAGCAGGTTCTGTACTTGGTAAT 237
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 238 GCTTTAAATAAAGGAATGGCTCCGCAAGCTAACTTAGTCTTCCAATCTATTATGATAGC 297
 QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 298 ACGCGAGGATTAGTGGCTTACCATCGAACTTAATACGTTATTATTAGTCAAGCTTGGAA 357
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTTPGlyAlaAlaValAsnGlyAlaTyrThrThr 140
 Db 358 GCTGGAGCAAGAATTCATACACTCTTTGGGAGCCCGAGTAAATGGAGCGTACACTGCT 417
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 418 AACTCGAGCAAGTGGATGATATGTTGAAATAATGATATGACGCTACTTTTTCACACT 477
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 478 GGTAAATGAAGGTCCTAAATTCAGGAACAATTAGTGTCCAGGTACAGCGAATAATGCTATT 537
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAsnAsnIleAsn 200
 Db 538 ACGGTGCGCGCAACCGAAACCTATCGCCACAGCTTCGGTTCGATAGCAGATAACCAAT 597
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 598 CATATTGCACAAATTTTCATCGAGAGGAGCTACGAGGATGACGAATTAAGCCTGACGTA 657
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAsnSerSerPhe 240
 Db 658 ACAGCTCTCTGGAACTTTATTTTATCAGCAGCTGTTCTTCTTAGCTCCAGACTCTTCGTT 717
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 718 TGGCGCAATTATAACAGTAATACCGGTATATGGCGGTACTCTCATGCGCACACTATT 777
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 778 GTTGCAAGGAATGTCCGCAATTAGCTGAGCATTTTATAAAAAATAGAGGTATTACTCTCT 837

```
QY 281 LysProSerLeuLeuLysAlaAlaLeuLleAlaGlyAlaAlaAspLleGlyLeuGlyTyr 300
Db 838 AAGCCTCTCTTTAATAAAGCTGACCTATCGTGGTGTCTACTGATGTTGGTTTAGGATAT 897
QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 898 CCTAATGGTGACCAAGCTGGGGCGTGTACTCTAAATAAATCGTTAAATGTAGCGTAT 957
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 958 GTCAATGAAGCAACTGCATTAGCCACAGGACAAAAGCAACGTAATTCGTTCCAGCACAA 1017
QY 341 AlaGlyLysProLeuLysLysLeuValTyrSerAspAlaProAlaSerThrThrAla 360
Db 1018 GCGGTAACCTTTAAATCTCGTAGTAGGACAGAGTCTCTCGGAAGTACAACTGCA 1077
QY 361 SerValThrLeuValAsnAspLeuAspLeuValLleThrAlaProAsnGlyThrGlnTyr 380
Db 1078 TCTTATACACTAGTTAATGATTTAGATCTAGTTATTACTGCTCCGAATGGACAAATAT 1137
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnValGlu 400
Db 1138 GTAGGAAATGATTTTGTAGTTATCTTATGATATACTGGGATGGTGGCAACATGTTGAG 1197
QY 401 AsnValPheLleAsnAlaProGlnSerGlyThrTyrThrLleGluValGlnAlaTyrAsn 420
Db 1198 AAGCTATTTATAAAGCTCCGAATCTGGAACGTAATATAATTTGAGGTTCAAGCGTATAAT 1257
QY 421 ValProValGlyProGlnThrPheSerLeuAlaLleValAsn 434
Db 1258 GTACCATCTGGCCACAGCGTTTCTCAGTAGTATCGTACAT 1299

RESULT 14
E03808
LOCUS      2218 bp      DNA      linear      PAT 29-SEP-1997
DEFINITION DNA encoding alkaline protease Ya.
ACCESSION E03808
VERSION    E03808.1 GI:2172022
KEYWORDS   JP 1992197182-A/1.
SOURCE     Bacillus sp.
ORGANISM   Bacillus sp.
REFERENCE  1. (bases 1 to 2218)
AUTHORS    Tobe S., Odera M. and Asai Y.
TITLE      DNA CODING ALKALINE PROTEASE YA ENZYME AND PRODUCTION OF ALKALINE
           PROTEASE YA USING THE DNA
JOURNAL    Patent: JP 1992197182-A 1 16-JUL-1992;
           LION CORP
COMMENT    OS Bacillus sp.
           PN JP 1992197182-A/1
           PD 16-JUL-1992
           PF 28-NOV-1990 JP 1990327110
           PI TOBE SEIICHI, ODERA MOTOYASU, ASAI YOSHIO
           PC C12N15/57,C12D3/386,C12N9/54,(C12N15/57,C12R1.07),(C12N9/54,
           PC C12R1.07);
           CC strandedness: Double;
           CC topology: Linear;
           CC hypothetical: No;
           CC anti-sense: No;
           CC *source: strain=y strain;
           CC *source: clone=pUB8A;
           FH Key      Location/Qualifiers
           PH CDS      218..2125
           FT          /product='precursor of alkaline protease Ya'
           FT sig_peptide 218..823
           FT mat_peptide 824..2122
           FT          /product='Precursor of alkaline protease Ya'.
           FT          Location/Qualifiers
           FT          1..2218
           FT          /organism="Bacillus sp."
           FT          /mol_type="genomic DNA"
           FT          /db_xref="taxon:1409"
```

```
ORIGIN
Alignment Scores:
Pred. No.:      1,01e-116      Length:      2218
Score:          1986.50        Matches:    379
Percent Similarity: 94.01%     Conservative: 29
Best Local Similarity: 87.33%   Mismatches: 25
Query Match:    88.41%        Indels:     1
Db:             6              Gaps:       1

US-09-985-689A-1 (1-434) x E03808 (1-2218)
QY 1 AsnAspValAlaAarGgGlyVileValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 824 AATGATGTAGCAGAGGGATAGTAAAGCTGATGTTGCACAAACAATATACGGATTATAT 883
QY 21 GlyGlnGlyGlnLleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 884 GGACAAGGTCAACTAGTTGCGTAGCGGACACAGGCTTAGATACAGGTCGTAAACGATAGT 943
QY 41 SerMethHisGluAlaPheArgGlyLysLleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 944 TCTATGATGATGAGCATTTCCGCGGGAATAATCAGAGCTCTTTACCGGTAGGAGAACTAAT 1003
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 1004 AATCGAGTGATCCGAATGGCATGGCACATGTAGCAGGTTCTGTACTTGGTAAT--- 1060
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerLleMetAspSer 100
Db 1061 GCITTAATAAAGGATGGCTCCGCAAGCTAACTTAGTCTTCCACTATTATATGGATAGC 1120
QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 1121 AGCGGAGGATTAGTGCTTTACCATCGAAGCTTAAATACGTTATTTAGTCAAGCTGGAAT 1180
QY 121 AlaGlyAlaArgLleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThr 140
Db 1181 GCTGAGCAAGAATTCTAATACTCTTGGGGAGCCCGAGTAATAGAGGCTACACTGCT 1240
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrLleLeuPheAlaAla 160
Db 1241 AACTCGAGACAAAGTGGATGATGTTGCAATAATATGATGATGACGCTACCTTTTTCAGCT 1300
QY 161 GlyAsnGlyProAsnGlyGlyThrLleSerAlaProGlyThrAlaLysAsnAlaLle 180
Db 1301 GGTAAATGAGGTCCTAATTCAGGAACAATTAGTCTCCAGGTACAGCAAAAATGCTATT 1360
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnLleAsn 200
Db 1361 ACGGTGCGGCGCACGGAAACTATCGCCCAAGCTTCGGTTCGATGACAGATAACCCAAAT 1420
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgLleLysProAspVal 220
Db 1421 CATATTGCACAAATTTTCATCGAGAGGAGCTACGAGGAGTACGAGGAGTAAAGCCTGACGTA 1480
QY 221 MetAlaProGlyThrPheLleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1481 ACAGCTCTGGAACATTTATTTATCAGCAGCTTCTTCTTCTAGCTCCAGACTCTTCGTTT 1540
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1541 TGGCGGAATTATAACAGTAATAACCGGTATATGGGGGTACCTCCATGCGCACACCTATT 1600
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLleThrPro 280
Db 1601 GTTCAGGGAATGCGCGCAATTTACGTGAGCATTTTATAAAAAATAGAGGTATTACTCCT 1660
QY 281 LysProSerLeuLeuLysAlaAlaLeuLleAlaGlyAlaAlaAspLleGlyLeuGlyTyr 300
Db 1661 AAGCCTCTCTTTAATAAAGCTGCACCTATCGTGGTGTCTACTGATGTTGGTTTAGGATAT 1720
QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerSerLeuAsnValAlaTyr 320
```



```
Qy 258 ThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGly 277
Db 16558 ACAGCATAGCAGGTGTTTTCAGATCCCTCGACGTGAATATCTCCTGAAGTGTCTGGA 16599
Qy 278 IleThr---ProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIle 296
Db 16598 ATCAACAACACCATCTGCAAGCCTCATTAAGCAACTCTCATCAACGGGGCAGAAGATTTA 16539
Qy 297 GlyLeu---GlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSer 315
Db 16538 GGCACACGGACATTCCTTAACGNAACAGGAGATGGGGGCAAAATGATCTCGAAATATAGC 16479
Qy 316 LeuAsnValAla-----TyrValAsnGluSerSerSerLeu 327
Db 16478 CTGAATCTCTCGTCTGCTGGGTTTCTCTGATGTCCTTCAGGATGATGAGCGTGAACCTT 16419
Qy 328 SerThrSerGlnLysAlaThrTyrSerPheThrAlaThrAlaGlyLysProLeuLysIle 347
Db 16418 CAGCAGGATTCTCATTTGATTTATCTCTTTGACCTTGATGGATCAAGGGTATTGACATT 16359
Qy 348 SerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThr-----Leu 364
Db 16358 ACCTTGGCATGGACGGATGCAGAGCCAGTGCAAATGCTGCACAACTGTAATCAGCTCTA 16299
Qy 365 ValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAsp 384
Db 16298 TTGAACAATTAGATTGATTTGATTTCTAAATTGCACCCGATGGGTCTTCATACCTTGGTAAATGAC 16239
Qy 385 PheThrSerProTyrAsn-----AspAsnTyrAspGlyArgAsnAsnValGluAsn 401
Db 16238 TTCTCTTCTGGGATTTCAACACACAGCGGTTCCGCTGATTAATCTGAACATATCGAGCGC 16179
Qy 402 ValPheIleAsnAla-----ProGlnSerGlyThrTyrThrIleGluVal----- 416
Db 16178 ATTCGCATCCCTGCAGGAGCGACGACACAGAACGGTGTGATGGATGGTTACAGTTGAACAT 16119
Qy 417 -----GlnAlaTyrAsnValProVal 423
Db 16118 CGAGGAGGAAGTTCTCAGCGCTACAGTATGTTATT 16083
```

Search completed: March 16, 2004, 00:36:45
Job time : 3658 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 22:28:03 ; Search time 396 Seconds
(without alignments)
4655.853 Million cell updates/sec

Title: US-09-985-689a-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAVNPVGPOTFSLAIV 434

Scoring table:
BLOSUM62 Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-C=/cgn2/1/USPTO_epool/US0985689/runat_10032004_112902_19707/app_query.fasta_1.583
-DB=N Geneseq_29Jan04 -OFW=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFW=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRF=US0985689 @CGN.1 1 352 @runat_10032004_112902_19707 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=10 -YGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2247	100.0	1923	2 AAX37278	Aax37278 Bacillus
2	2242	99.8	1923	2 AAX37279	Aax37279 Bacillus
3	2183	97.2	1920	2 AAX37277	Aax37277 Bacillus
4	2125.5	94.6	3003	2 AAV82382	Aav82382 Bacillus
5	1986.5	88.4	1299	2 AAQ27516	Aaq27516 Alkali-pr
6	452.5	20.1	1977	2 AAT85667	Aat85667 Thermococ
7	452.5	20.1	1977	2 AAX05926	Aax05926 WO9856926
8	416.5	18.5	1236	2 AAX05920	Aax05920 Hyperther

RESULT 1
AAX37278

ID AAX37278 standard; DNA; 1923 BP.

XX AAX37278;
AC
XX
XX
DT 20-MAR-2003 (revised)
DT 21-JUL-1993 (first entry)
XX
XX
DE DE Bacillus alkaline protease encoding DNA.
XX
XX Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
KW washing composition; oxidising agent; ss.
XX
XX Bacillus sp.
XX
XX WO9818218-Al.
XX
XX
PD 15-APR-1999.
XX
XX 07-OCT-1998; 98WO-JP004528.
XX
XX 07-OCT-1997; 97JP-00274570.
XX
XX (KAOS) KAO CORP.
XX
XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
PI Shikata S, Nomura M;
XX
XX Aaq27516 Alkali-pr
DR WPI; 1999-287736/27.
XX
XX P-PSDB; AAY17088, AAY17090.

ALIGNMENTS

9	416.5	18.5	1566	2 AAT85668	Aat85668 Pyrococcus
10	416.5	18.5	1962	2 AAT85695	Aat85695 Pyrococcus
11	416.5	18.5	1962	2 AAX05929	Aax05929 Hyperther
12	403.5	18.0	1977	2 AAT85669	Aat85669 Protease
13	372	16.6	2121	4 ABL54900	Ab154900 T. yonsei
14	346	15.4	2539	2 AAT61454	Aat61454 Streptomy
15	346	15.4	2809	2 AAT61455	Aat61455 DhpA-mel
16	338	15.0	135638	7 ABX34289	Abx34289 S. atrool
17	332.5	14.8	1329	6 ABK74643	Aak74643 Bacillus
18	310.5	13.8	2532	2 AAQ29134	Aaq29134 Encodes R
19	307.5	13.7	1859	2 AAT85677	Aat85677 Thermococ
20	307	13.7	2835	2 AAT08141	Aat08141 Hyperther
21	307	13.7	4765	2 AAT08132	Aat08132 protease
22	307	13.7	4765	2 AAT85670	Aat85670 Pyrococcus
23	307	13.7	4765	2 AAX05921	Aax05921 WO9856926
24	306.5	13.6	1560	7 ABZ37569	Abz37569 Streptomy
25	306.5	13.6	59816	7 ABZ37516	Abz37516 Streptomy
26	306.5	13.6	59816	7 ABZ37515	Abz37515 Streptomy
27	301.5	13.4	3413	2 AAV72330	Aav72330 F. balust
28	296	13.2	898	2 AAT08131	Aat08131 Hyperther
29	290.5	12.9	564	2 AAT08134	Aat08134 DNA seque
30	290.5	12.9	564	2 AAT85676	Aat85676 Thermococ
31	283	12.6	3743	9 ADD24901	Add24901 DNA encod
32	282	12.6	3788	9 ADD24905	Add24905 DNA encod
33	276	12.3	1306	6 ABL55784	Ab155784 Bacillus
34	276	12.3	1330	6 ABL55787	Ab155787 Bacillus
35	275	12.2	546	4 ABL53453	Ab153453 T. yonsei
36	265.5	11.8	2273	2 AAO04339	Aaq04339 Aquaricin
37	265.5	11.8	2274	2 AAO12838	Aaq12838 Aquaricin
38	265.5	11.8	2274	2 AAO75859	Aaq75859 Aqualysin
39	263	11.7	4716	9 ABQ80437	Abq80437 Cspa codi
40	263	11.7	4740	7 ABZ58957	Abz58957 Group B S
41	256.5	11.4	1152	8 ACC85473	Acc85473 B gibsoni
42	256.5	11.4	1194	8 ADA32120	Ada32120 DNA encod
43	256.5	11.4	10216	2 AAT39279	Aat39279 Transposo
44	256	11.4	1485	6 ABK74647	Abk74647 Bacillus
45	256	11.4	4650	6 ABN71526	Abn71526 Streptoco

PT Alkali protease from *Bacillus* used in washing powders.

XX PS Disclosure; Page 58-63; 71pp; Japanese.

XX The invention relates to alkaline proteases produced by strains of
CC *Bacillus*. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidising agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC stable over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
CC used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease encoding
CC DNA. (Updated on 20-MAR-2003 to correct DR field.)

XX SQ Sequence 1923 BP; 576 A; 418 C; 475 G; 454 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.13e-163 Length: 1923
Score: 2247.00 Matches: 434
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-09-985-689A-1 (1-434) x AAX37278 (1-1923)

QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnSerTyrGlyLeuTyr 20
DB 619 AATGATGTTGGCGTGGATTGTCAAAGCGGATGTGGCTCAGACGAGTACGGGTGTAT 678
QY 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 679 GGAACAAGGACAGATCGTCGGTTCGGGATACAGGCTGTGATACAGGTCGCAATGACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 739 TCGATCATGAGCTTCGCGGGGAAATTAAGTGCATTATATGATTCGGACGCGAAT 798
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 799 AATGCCAATGATACGAATGCTCATGTGTCATGTCATGTCGCTGGCTCCGCTATTAGGAACGCG 858
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 859 TCCACTAATAAGGAATAGCGCTCAGGCGAATCTAGTCTTCCATCTATCATGATAGC 918
QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
DB 919 GGTGGGGGACTTGGAGGACTTACCTCGAATCTGCAACCTTATTTCAGCCCAAGCATACAGT 978
QY 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyValTyrThr 140
DB 979 GCTGGTCCGAGATTTCATCAAACTCTCGGAGCGACGATGAATGGGCTTTCACAAACA 1038
QY 141 AspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 1039 GATTCAGAAATGTGATGACTATGTGCGCAAAATGATATGACGATCTCTTCGCTGCC 1098
QY 161 GlyAsnGlyGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
DB 1099 GGGAAATGAAGGACCGGAACCGCGGAACCATCAGTGCACCGGACAGCTAAATAATGCAATA 1158
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
DB 1159 ACAGTCGAGCTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATCGGACATATCAAC 1218
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220

DB 1219 CATGTGGCACATTCTCTTCACGTGGACCGCAAGGATGGAGGATCAACCGGATGTC 1278
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
DB 1279 ATGGCACCGGAAAGTTCATCTATCAGCAAGATCTCTCTTGCACCGGATCTCTCTTC 1338
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
DB 1339 TGGCGCAACCATGACAGTAAATATGATACATATGGGGAACGTCCTATGGCTACCGCATC 1398
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
DB 1399 GTTGTGGAAACGTGGCACAGCTTCGTGAGCATTTTGTGAAAAACACAGAGCATCACACCA 1458
QY 281 LysProSerLeuLeuLysAlaAlaIleAlaGlyAlaAspIleGlyLeuGlyTyr 300
DB 1459 AAGCTTCTCTATAAAGCGGCACGTATGCGGTGACGCTGACATCGGCTTGGCTAC 1518
QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
DB 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATTGGATAAATCCCTGAACGTTGCTAT 1578
QY 321 ValAsnGlnSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
DB 1579 GTGACGAGTCCAGTTCCTCTATCCACGACCAAAAGCGACGTACTCGTTTACTGTACT 1638
QY 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAla 360
DB 1639 GCGGCAAGCCTTGAANAATCTCCCTGGTATGCTGATGCCCTCGGAGCACACTGCT 1698
QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
DB 1699 TCGTAACGCTTGTCAATGATCTGGACCTTGTCATTCGCTCCAAATGGCACACAGTAT 1758
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnValGlu 400
DB 1759 GTAGGAATGACCTTACTTCCCATACATGATGAATGGATGGCGGCAATACGTAGAA 1818
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
DB 1819 AATGATATTATTAATGACCAACCAAGCGGACGCTATACAAATTGAGGTACAGGCTTATAAC 1878
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
DB 1879 GTACCGTTGGACACACAGCTTCTGTTGGCAATTGTGAAT 1920
RESULT 2
AAX37278
ID AAX37279 standard; DNA; 1923 BP.
XX AC AAX37279;
XX XX
DT 20-MAR-2003 (revised)
DT 21-JUL-1999 (first entry)
XX XX
DE *Bacillus* alkaline protease encoding DNA.
XX KW Alkaline protease; *Bacillus*; casein digestion; oleic acid; enzyme;
KW washing composition; oxidising agent; ss.
XX OS *Bacillus* sp.
XX PN WO9918218-A1.
XX PD 15-APR-1999.
XX PF 07-OCT-1998; 98WO-JP004528.
XX PR 07-OCT-1997; 97JP-00274570.
XX PA (KAO) KAO CORP.
XX PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;

PI Shikata S, Nomura M;
 XX WPI; 1999-287736/27.
 DR P-PSDB; AAY17091.
 XX
 PT Alkali protease from *Bacillus* used in washing powders.
 XX
 PS Disclosure; Page 63-68; 71pp; Japanese.
 XX
 CC The invention relates to alkaline proteases produced by strains of
 CC *Bacillus*. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 1923 BP; 578 A; 417 C; 474 G; 454 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2,75e-163 Length: 1923
 Score: 2242.00 Matches: 433
 Percent Similarity: 99.77% Conservative: 0
 Best Local Similarity: 99.77% Mismatches: 1
 Query Match: 99.78% Indels: 0
 DB: 2 Gaps: 0
 US-09-985-689A-1 (1-434) x AAX37279 (1-1923)
 QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnSerTyrGlyLeuTyr 20
 DB 619 AATGATGTTGGGTGGGATGTTGCAAGCGGATGTGCTCAGACGAGCTACGGGTGTAT 678
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 679 GGACAGGACAGATCGTAGCGGTGCGATACAGGCGTTGATACAGGTGCGAATGACAGT 738
 QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 DB 739 TCGATGATGAAGCTTCCCGGGGAAATACGCAATATATGATGATGGACGCGAAT 798
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 DB 799 AATGCCAATGATACGATGTCATGTACGATGTGCTGGCTCCGTATTAGGAAACGGC 858
 QY 81 SerThrAsnIysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 DB 859 TCCACTAATAAGGAATGGCGCTCAGCGCAATCTAGTCTTCCAATCTATCATGGATAGC 918
 QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 DB 919 GGTGGGGAGCTTGAGGAGCTACCTTCGAATCTGCAACCTTATTCAGCAGGATACAGT 978
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
 DB 979 GCTGGTGCAGAAATTCATAAACTCTCTGGGGAGCAGCAGTGAATGGGGCTTACACAACA 1038
 QY 141 AspSerArgAsnValAspAspTyrValArgIysAsnAspMetThrIleLeuPheAlaAla 160
 DB 1039 GATTCCAGAAATGGGATGACTATGTGGCAAAATGATATGACATCTCTTTTCGTGCC 1098
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaIysAsnAlaIle 180
 DB 1099 GCGAATGAAGACCGAAGCGGCGGAACCATCAGTCACCGCAGCAGTAAATAATGCAATA 1158
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200

DB 1159 ACAGTCGGAGCTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATCGGCAATATCAAC 1218
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrIysAspGlyArgIleIysProAspVal 220
 DB 1219 CATGTGGACAGATTCTCTTCACGTGGACCCGACAAAGGATGGACGATCAACACCGGATGC 1278
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 DB 1279 ATGGACCGGGAACGTTCACTATCATCAGCAAGATCTTCTCTTGCACCGGATCTCTCTTC 1338
 QY 241 TrpAlaHisAspSerIysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 DB 1339 TGGCGCAACCATGACAGATAATATGCATACATGGGTGGAACGTCATGGGTACACCGATC 1398
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyIleThrPro 280
 DB 1399 GTTCTGGAAACGTTGGCACAGCTTCGTGAGCATTTTGTGAAACACAGGCATCACACCA 1458
 QY 281 LysProSerLeuLeuIysAlaAlaLeuIleAlaGlyAlaAlaIleGlyLeuGlyTyr 300
 DB 1459 AAGCCTTCTCTATTAAAGCGGCACCTGATTGCCGCTGCAGCTGACATCGGCCTTGGGTAC 1518
 QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspIysSerLeuAsnValAlaTyr 320
 DB 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATGATGATTAATCCCTGACGTTGCCAT 1578
 QY 321 ValAsnGluSerSerLeuSerThrSerGlnIysAlaThrTyrSerPheThrAlaThr 340
 DB 1579 GTGAACGAGTCCAGTTCTCTATCCACAGCCAAAGCGACGTACTCGTTTACTGTACT 1638
 QY 341 AlaGlyIysProLeuIysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 DB 1639 GCCGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCCCTGGAGCAACACTGCT 1698
 QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 DB 1699 TCCGTAAACGCTTGTCAATGATCTCGACCTTGTCTATTACCGCTCCAAATGGCACAGTAT 1758
 QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 DB 1759 GTAGGAAATGACTTTACTTCGCCATACAATGATAACTGGGATGGCCCAATAACGTAGAA 1818
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 DB 1819 AATGTATTATTATATGACACCAAGCGGACGTATACAATTGAAGTACAGGCTTATAAC 1878
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 DB 1879 GTACCGGTTGGACCACACAGAACTTCTCGTTGGCAATTGTGAAT 1920
 RESULT 3
 AAX37277
 ID AAX37277 standard; DNA; 1920 BP.
 XX
 AC AAX37277;
 XX
 XX 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE *Bacillus* alkaline protease encoding DNA.
 XX
 KW Alkaline protease; *Bacillus*; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent; ss.
 XX
 OS *Bacillus* sp.
 XX
 PN WO9918218-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX

PR 07-OCT-1997; 97JP-00274570.
 XX (KAOS) KAO CORP.
 PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX WPI: 1999-287736/27.
 DR P-PSDB; AAV17087, AAV17089.
 XX Alkali protease from *Bacillus* used in washing powders.
 PT Disclosure; Page 53-58; 71pp; Japanese.
 XX The invention relates to alkaline proteases produced by strains of
 CC *Bacillus*. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 1920 BP; 576 A; 411 C; 473 G; 460 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 9,57e-159 Length: 1920
 Score: 2183.00 Matches: 418
 Percent Similarity: 99.31% Conservative: 13
 Best Local Similarity: 96.31% Mismatches: 3
 Query Match: 97.15% Indels: 0
 DB: 2 Gaps: 0

US-09-985-689A-1 (1-434) x AMX37277 (1-1920)

QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnSerTyrGlyLeuTyr 20
 Db 616 AATGATGTCCGACAGGATTTGTCAAAGCGGATGTGGCACAGCAGCTACGGTTTGTAT 675
 QY 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 676 GGACAAAGCCAGATTCTCCAGTTGCCGATCTGGATTCGATACAGGAAGAACACAGT 735
 QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 736 TCGATGCATCAAGCTTCGCGGTAAATAACAGCACTATATGCACTGGTCCGACGAAT 795
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 796 AATGCGAATGATACGAACGGTCTATGTTACCCATGTCAGGTTCGGTATTAGGAATGGC 855
 QY 81 SerThrAsnIscGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 856 GCACAGATTAAGGATGGACCTCAAGCGAATCTGGTTTTCAATCCATCATGATAGC 915
 QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 916 AGTGGTGGGCTTGAGGCTTCCCTCAATCTGCAAAACCTTATTCAGCCCAAGCATTCAGT 975
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValAlaValAsnGlyValTyrThrThr 140
 Db 976 GCAGGTCCCAAGATTATCAAACTCTGGGGGGCAGCGGTGATGGGCCCTACAGACA 1035
 QY 141 AspSerArgAsnValAspAspTyrValArgIysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1036 GATTCCAGAAATGTGGATGACTATGTAAGGAAATGATGACGATCTTTTCGGCGCT 1095

QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaIysAsnAlaIle 180
 Db 1096 GGGAAATGAAGCGCGGACCGGTCACCATCAGTCACCTGGTACGGCTAAACGCCATA 1155
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1156 ACAGTCGGCGCACACCGAAACCTGCTCAAGCTTCGGTTCCTATGACGATATATATAAC 1215
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyValArgIleIysProAspVal 220
 Db 1216 CACGTTGCACAGTTCCTCCGTCGCGCCGACAAAGATGGCGAATCAAGCTGATGTC 1275
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 1276 ATGGCGCCAGGACATACATTTTATCAGCAAGATCTCTCTGCACCCGATTCCTCTCTTC 1335
 QY 241 TrpAlaAsnHisAspSerLysIleValAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 1336 TGGCGCAATCAGACGCAATATGCTATATGGGTGGACGCTCCATGGCAACCGATT 1395
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyIleThrPro 280
 Db 1396 GTTGGCGGGAATCTGCACAGCTCCGTCGAGCATTTTGTGAAAAATAGAGGAATCACTCT 1455
 QY 281 LysProSerLeuLeuLysAlaAlaIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1456 AAGCTTCCCTATTGAAGCAGCTTTGATTGAGGTGCTGCTGATGTTGGATTGGGTTAT 1515
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1516 CCGAACGAAACCAAGGATGGGCGCGAGTGACCTGGATAAATCGTTGAACGTTGCCTAT 1575
 QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 1576 GTGAACGAATCCAGTGCCTTATCACTAGCCAAAAGACGACATATACCTTTACTGCAACG 1635
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 1636 GCGGCGAGCGCATTTGAAAATCTCCCTGGTATGTCGATGTCCTCCGACACTACTGCT 1695
 QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1696 TCTGTAACTCCCTGTCATGATTGGATTGGTTCATTACAGCACCAACGAAACAGATAT 1755
 QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 Db 1756 GTCGGGAATGACTCTCAGCACCATTTGCAATAACTGGGATGGCGCAATAACGTAAGAA 1815
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1816 AATGTATTTTAAATTCGCCCCAAAGTGGAAACATATACCATTTGAGGTGCAAGCATATAAT 1875
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1876 GTCCGGTTGGACCAAAAACCTTCCTGTTGGCAATTGTGAAC 1917

RESULT 4
 AAV82382 standard; DNA; 3003 BP.
 ID AAV82382
 XX AC AAV82382;
 XX DT 12-APR-1999 (first entry)
 XX DE *Bacillus* JP170 protease gene.
 XX KW Protease; detergent; surfactant; leather processing; debittering;
 XX KW flavour; ss.
 XX OS *Bacillus* sp.
 XX FH Key Location/Qualifiers
 CDS 846..2771

FT sig_peptide /*tag= a
 FT 846..944
 FT /*tag= b
 FT mat_peptide 1470..2768
 FT /*tag= c

XX WO9856927-A2.

XX 17-DEC-1998.

XX 09-JUN-1998; 98WO-US012005.

XX 12-JUN-1997; 97US-00873479.

XX (NOVO) NOVO NORDISK BIOTECH INC.

XX Sloma A, Christianson L;

XX WPI; 1999-080908/07.

DR P-PSDB; AAW89547.

XX Novel protease from *Bacillus subtilis* LC20 - useful in laundry and
 PT dishwashing detergents and for leather processing.

XX Claim 11; Page 52-53; 77pp; English.

CC This nucleotide sequence encodes a novel protease (see AAW82382) of
 CC *Bacillus* sp. JP170 (NCIB 12513). The sequence in plasmid p170BAN is
 CC contained in *Bacillus subtilis* LC20 NRRL B-21680. The protease gene was
 CC isolated from chromosomal DNA of JP170 following preparation of probes
 CC based on protease N-terminal and internal peptides (see AAW89549-50).
 CC screening of chromosomal libraries, isolation of the 3' end of the gene
 CC by inverse PCR (see AAW84410-11), reconstruction of 5' and 3' ends and
 CC PCR amplification (see AAW82412-16). Claimed recombinant host cells can
 CC be used in a method for producing the protease. The protease is used in
 CC laundry and dishwashing detergents, for institutional and industrial
 CC cleaning, and for leather processing, as well as for debittering and
 CC enhancing the degree of hydrolysis of protein hydrolysates, for flavour
 CC development through hydrolysis of proteins, degradation of undesired
 CC peptides and in enzymatic synthesis of peptides. It has enhanced
 CC stability towards oxidation under alkaline conditions, e.g. towards
 CC bleaching agents of the peroxy type. The invention also provides mutant
 CC cells in which the protease activity is diminished. Such cells can be
 CC used for the production of heterologous recombinant proteins

SQ Sequence 3003 BP; 982 A; 504 C; 645 G; 872 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4,35e-154 Length: 3003
 Score: 2125.50 Matches: 406
 Percent Similarity: 97.93% Conservative: 19
 Best Local Similarity: 93.55% Mismatches: 8
 Query Match: 94.59% Indels: 1
 DB: 2 Gaps: 1

US-09-985-689A-1 (1-434) x AAW82382 (1-3003)

Qy 1 AsnAspValAlaArgGlyIleValIleAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 1470 AATGACGTGGCCGTGGCATTTGGTGAAGCAGACGTCGCACAAATAACITTTGGCTTATAT 1529
 Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 1530 GGACAAGGACAGATTGTAGCAGTTGCTGATCTGGGCTTGATACAGGAAGAAATGACAGT 1589
 Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 1590 TCGATGTCATGAGCATTTCCCGGTCAGATTACCGCACTATATGCCTGGGCAGAACGAAT 1649
 Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 1650 AACGCCAATGATCCAAATGGACATGGAAACCCATGTTCTGTCGATCTGTGTAGGAAT--- 1706

Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 1707 GCTACAAATAAAGGATGGCAGCGCAACTCTAGTCTCTTCAATCTATTATGGATAGT 1766
 Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 1767 GGTGGAGGGCTGGAGGACTACTCTCTAATCTCAACAACATTATTTCAGTCAAGCATATAGT 1826
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThrThr 140
 Db 1827 GCTGGAGCGAGAAATTCATACGAATTCATGGGGGCTCCAGTAAACGGTGCCTATACGACA 1886
 Qy 141 AspSerArgAsnValAspAspTyrValAlaGlyAsnAspMetThrIleLeuPheAlaAla 160
 Db 1887 GACTCTCGAAATGTTGATGATTATGTGAGAAATAATGATAGCAGATCTCTTTTGGGCC 1946
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1947 GGAAATGAGGGACCCAGCTAGCGGTACAATCAGTGCACCGAACAGCAAAATATGCGATT 2006
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 2007 ACAGTTGGGCAACCGAAACCTACGTCACAGCTTCGGATCTTATGCGGATATATTAAC 2066
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 2067 CATGTTGCTCAATTCCTTCAAGAGTCTCTACAGATGAGATGAGCTATTAAAGCGGACGTC 2126
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 2127 ATGGCACCAGGTAGCTATATCTCTCTGCTAGATCATCTTAGCTCCAGATTCCTCATTC 2186
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 Db 2197 TGGCAAAACCATGATAGTAATATGCTACATGGGTGCTACTTATGGCTACTCCATT 2246
 Qy 251 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 2247 GTAGCAGGTAATGTTGCACAAATTAAGGAGCATTTTGTGAAAAATAGAGGGTAACCTCT 2306
 Qy 281 LysProSerLeuLeuValAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 2307 AAGCTTCCCTTTTAAAGCTGCTTTTAAATGTCAGTCTCTGGGATGTTGGACTTGGCTTT 2366
 Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 2367 CCAATGCTAAACCAAGGATGGGAAGAGTAACGTTAGATAAATCCCTAAATGTGCGATT 2426
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 2427 GTGAATGAACGAGCGCCCTTTATCAACAGTCAAAAGCAACATATTCTTTACGGCTCAA 2486
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrAla 360
 Db 2487 GCTGGTAAACCTTTAAAAATATCACTTTGTTGGTCAGATGCACCGTAGCAGCGCA 2546
 Qy 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlyTyr 380
 Db 2547 TCACCTAACTTTAGTGAATGATTAGACTTAGTAATCACTGCACCAATGGGAACATAATAC 2606
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnValGlu 400
 Db 2607 GTCGGAATGACTTTACAGCACCGTATGATAACAATTTGGGATGCGCAGAAACAACGTGAA 2666
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 2667 AATGTGTTTATCAATGCTCTCAAGCGGAACGTATACAGTCGAAGTGCAGGCTTACAAT 2726
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 2727 GTACCAATGATGCCGAAACCTTTTCTTAGCGATTGTACAT 2768

RESULT 5

424

AAQ27516
 ID AAQ27516 standard; DNA; 1299 BP.
 XX
 AC
 AAQ27516;
 XX
 DT 05-FEB-1993 (first entry)
 XX
 DE Alkali-protease Ya enzyme gene.
 XX
 KW Alkali resistance; surface active agent resistance; detergency improver;
 KW ss.
 XX
 OS Bacillus sp. Y.
 XX
 FH Key Location/Qualifiers
 CDS 1..1299
 FT /*tag= a
 FT
 XX JP04197182-A.
 XX
 PD 16-JUL-1992.
 XX
 PF 28-NOV-1990; 90JP-00327110.
 XX
 PR 28-NOV-1990; 90JP-00327110.
 XX
 PA (LIOY) LION CORP.
 XX
 DR WPI; 1992-288440/35.
 DR P-PSDB; AAR26274.
 XX
 PT DNA coding alkali-protease Ya enzyme - has good alkali and surfactant
 PT resistance and improves detergency.
 PS Claim 3; Page 2; 17pp; Japanese.
 XX
 CC The sequence is that of the alkali-protease Ya enzyme gene which can be
 CC used in the recombinant production of ya enzyme. ya enzyme is excellent
 CC in alkali resistance and surface active agent resistance and improves
 CC detergency
 SQ Sequence 1299 BP; 401 A; 246 C; 297 G; 355 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 8,12e-144 Length: 1299
 Score: 1986.50 Matches: 379
 Percent Similarity: 94.01% Conservative: 29
 Best Local Similarity: 87.33% Mismatches: 25
 Query Match: 88.41% Indels: 1
 DB: Gaps: 1

US-09-985-689A-1 (1-434) x AAQ27516 (1-1299)

QY 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnSerTyrGlyLeuTyr 20
 Db 1 AATGATGTAGCAAGAGGATAGTAAAGCTGATGTGCACAAAACAATATACGATTATAT 60
 QY 21 GlyGlnGlyGlnLeuValAlaValAlaAspThrGlyLeuAspThrGlyArgAspSer 40
 Db 61 GGACAGGTCACTAGTTCAGTAGCGGACACAGGCTTAGACAGTCTGTAACGATAGT 120
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGATGAAGCATTCGCGGGAATAACACAGCTCTTTACGCGTTAGGAAGAACTAAT 180
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValleuGlyAsnGly 80
 Db 181 AATGCGAGTATCCGAATGGCATGGCAGACACATGTAGCAGGTCTGTACTTCGTAAT--- 237
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 238 GCCTTAATAAAGGAATGGCTCCGCAAGCTAACTTAGTCTTCCAATCTATTATGATAGC 297

QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 298 AGCGGAGGATTAGTGGCTTACCATCGAACTTAATACTGTTATTATTAGTCAGCTTGAAT 357
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
 Db 358 GCTGGAGCAAGAATTCACTAACTCTTTGGGAGGCCCAAGTAAATGGAGCGTACACTGCT 417
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 418 AACTCGAGCAAGTGGATGATGTTTCGAAATATGATATGACCGTACTTTTTCGAGCT 477
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 478 GGTATATGAAGTCTTAATTCAGGAACAATTAGTGTCCAGGTACACGCAAAATGCTATT 537
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 538 ACGTCGCGCGCAACGGAACAACTATCCCAAGCTTCGTTTCGATAGCAGTAACCCAAAT 597
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 598 CATATTGCACAATTTTCATCGAGAGGAGCTACGAGGATGGACGAATTAAGCTGACGTA 657
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 658 ACAGCTCCTGGAACATTTATTTATCAGCACGTTCTTCTTACCTCCAGCTTCCTCGTT 717
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 718 TGGCGCAATTAACAGTAATAACGCGTATATGGCGGTACCTCCATGCGCACACCTAAT 777
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 778 GTTCGAGGGAATTCGCGCAATTAACGCGTATTTTATAAAAAATAGAGGTATTACTCCT 837
 QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 838 AAGCTTCTTTATAAAGCTGACATTATCGCTGGTGCTACTGATGTTGTTTAGGATAT 897
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 898 CCTAGTGTGACCAAGCTGGGCGCTGTTACTCTAGATAAATCGTTAAATGTAGCGTAT 957
 QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 958 GTCAATGAGCAACTGATTTAGCCACAGACAAAGAACACGTTATTCGTCCAGACACAA 1017
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 1018 GCGGGTAAACCTTTAAAAATCTCGTTAGTAGACAGATGCTCCTCGAAGTACAACTGCA 1077
 QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1078 TCTTATACACTAGTAAATGATTAGATTCTAGTTATTACTGCTCCGAATGGACAAAAATAT 1137
 QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 Db 1138 GTAGGAATGATTTTAGTTATCTCTTATGATAATACTGGGATGCTCGCAACATGTTGAG 1197
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1198 AACGTATTTATAACGCTCCGCAATCTGGAACTGATATATTAATTGAGGTTCAAGCGTAAAT 1257
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1258 GTACCATCTGCCCAACAGCGTTTCTCCTACCTAGCTATCGTACAT 1299

RESULT 6
 AAT85667
 ID AAT85667 standard; DNA; 1977 BP.
 XX
 AC AAT85667;

XX 17-OCT-2003 (revised)
 DT 20-APR-1998 (first entry)
 XX
 DE Thermococcus protease coding sequence.
 XX
 KW Protease; research reagent; thermal stability; thermococcus celer; ss.
 XX
 OS Thermococcus celer; DSM-2476.
 XX
 XX WO9721823-A1.
 XX
 XX 07-NOV-1996; 96WO-JP003253.
 XX
 XX 12-DEC-1995; 95JP-00323285.
 XX
 XX (TAKI) TAKARA SHUZO CO LTD.
 XX
 XX Takakura H, Morishita M, Yamamoto K, Mita M, Asada K;
 PI Tsunawawa S, Kato I;
 PI
 XX WPI; 1997-332794/30.
 DR P-PSDB; AAW24121.
 XX
 XX Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT industrially and as research reagents.
 XX
 XX Claim 3; Page 86-87; 159pp; Japanese.
 XX
 XX This sequence represents the coding sequence for the protease from
 CC Thermococcus celer DSM-2476. This sequence encodes a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries. (Updated on 17-
 CC OCT-2003 to standardise OS field)
 XX
 SQ Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,72e-25 Length: 1977
 Score: 452.50 Matches: 138
 Percent Similarity: 44.66% Conservative: 67
 Best Local Similarity: 30.07% Mismatches: 153
 Query Match: 20.14% Indels: 101
 DB: 2 Gaps: 18

US-09-985-689A-1 (1-434) x AAT85667 (1-1977)

QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnLeuValAla 27
 DB 433 ATAGGGGCGGATACCGTCTGGAACTCCCTCGGCTACGACGGAAGCGGTGGTGGTGGC 492
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
 DB 493 ATCTGCGATACGGGTATAGACGCGAAC-----CACCCGATCTGAG 534
 QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
 DB 535 GCGAAGGTATAGGGTGGTACGACGCGCTCAACGCGCAGGTGACCCCTACGATGACCCAG 594
 QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
 DB 595 GGACACGGAAACCCAGCTTCGGGGTATCGTTGCCGGAACCGGAGCGGTAACTCCAGTAC 654
 QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
 DB 655 ATAGGGGTGCGCCCGGCGGAAGTCTGTCGCGCTCAAGGTCTCGTGGCGGACGCGTTCG 714
 QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121

RESULT 7

AA05926

ID AA05926 standard; DNA; 1977 BP.

DB 715 GGAAGCGTCTCCACCATCATCGCGGGTGTGACTGGGTCTCCAGAACACGACAAAGTAC 774
 QY 122 GlyAlaArgIle-----HisThrAsnSer 129
 DB 775 GGGATAAGGGTCTCAACACCTCTCCCTCGGCTCTCCAGAGCTCCGACGGAACCGACTCC 834
 QY 130 TrpGlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrVal 149
 DB 835 CTCAGTCAGCGCTCAACACGCTGGGACGCC----- 867
 QY 150 ArgLysAsnAspMetThrIleLeuPheAlaGlyAsnGluGlyProAsnGlyGlyThr 169
 DB 868 -----GGTATAGTAGTCTCGCTCGCCGCAACAGCGGCCCAACACCTACACCC 918
 QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
 DB 919 GTGCGCTACCCCGCGCGAGCAAGGTCAATACCGTCGGTCA----- 963
 QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
 DB 964 -----GTTGACAGCAACGACACATCGCCAGCTTCTCCAGCAGGGA 1005
 QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
 DB 1006 CCGACCGCGACGGAAGGCTCAAGCGGAAGTCTGTCGCCCGCGGCTTGACATCATGCC 1065
 QY 230 AlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyrAla 249
 DB 1066 CCGCGCGCCAGC-----GGAACCAAGCATGGCGACCCCGCATAAACGACTACTACACC 1116
 QY 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu 268
 DB 1117 AAGGCTCTGGAAACGACGATGCGCCCGACGTTTCGGGCTGGCGGCTCATCTC 1176
 QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
 DB 1177 CAGGCGCCAC-----CCGAGCTGAGCCCGGACAAAGGTGAAG 1212
 QY 287 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
 DB 1213 ACGGCCCTCATCGACACCGCCGACATAGTCGCCCAAGGAGATAGCGGACATCGCTTAC 1272
 QY 299 GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- 316
 DB 1273 GGTGCG-----GGTAGGGTGAAGCTCTACAAGGCCATCAAGTAC 1311
 QY 317 ---AsnValAlaTyrValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyr 335
 DB 1312 GACGACTACGCCAAGCTCACCTTCACCGCTCCGTCGCGACAAAGGAGCGCCACCCAC 1371
 QY 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
 DB 1372 ACCTTCGAGCTAGCGCGCCACCTTCGTGACGCGCCACCTCTACTGGGAC----- 1422
 QY 356 AlaSerThrThrAlaSerValThrLeuValAsnAlaPheLeuAspLeuValIleThrAlaPro 375
 DB 1423 -----ACGGGCTCGAGCGACATCGACCTCTACTCTACGACCCC 1461
 QY 376 AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGly 395
 DB 1462 AAGCGGAACGAG---GTTGACTACTCTACACGCGCTACTAC----- 1500
 QY 396 ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGlu 415
 DB 1501 -----GGCTTCGAGAAGGTTCGCTACTACAACCGACCGCGGAACCTGGACGGTCAAG 1554
 QY 416 ValGlnAlaTyrAsnValProValGlyProGlnThrPheSerIleAlaIleValAsn 434
 DB 1555 GTCTGACGTACAAG-----GGCGGCGGAACACTACAGGTGACGTGCTGACG 1602

XX AC AAX05926;
 XX DT 06-MAY-1999 (first entry)
 XX DE WO9856926 Seq ID 11.
 XX KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 XX KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
 XX OS Thermococcus celery.
 XX PN WO9856926-A1.
 XX PD 17-DEC-1998.
 XX PF 04-JUN-1998; 98WO-JP002465.
 XX PR 10-JUN-1997; 97JP-00151969.
 XX PA (TAKI) TAKARA SHUZO CO LTD.
 XX PI Takakura H, Morishita M, Shimojo T, Aeada K, Kato I;
 XX DR WPI; 1999-080907/07.
 XX DR P-PSDB; AAW94840.
 XX PT Recombinant hyperthermostable protease from *Pyrococcus furiosus* - and
 XX PT gene encoding it, for large scale production of the protease for
 XX PT industrial use.
 XX PS Disclosure; Page 53-54; 82pp; Japanese.
 XX CC The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
 CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially *Bacillus* strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis
 XX SQ Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.72e-25 Length: 1977
 Score: 452.50 Matches: 138
 Percent Similarity: 44.66% Conservative: 67
 Best Local Similarity: 30.07% Mismatches: 153
 Query Match: 20.14% Indels: 101
 DB: 2 Gaps: 18

US-09-985-689a-1 (1-434) x AAX05926 (1-1977)

QY 8 ValIysAlaAspValAlaGlnSerSerTyrrGlyLeuTyrrGlyGlnGlyGlnIleValAla 27
 Db 433 ATAGGGCCGATACCGTCTCGAATCTCCCTCGGTACGACGGAAGCGGTGGTGGTGGC 492
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
 Db 493 ATCGTCGATACGGGTATAGACGGGAC-----CACCCCGATCTGAAG 534
 QY 48 GlyIysIleThrAlaLeuTyrr-----AlaLeuGlyArgThrAsnAsnAlaAspThrAsn 66
 Db 535 GGCAGGTCTAGCGTGGTACGACGCGGTCAACGGCAGGTCGACCCCTACGATACGAC 594
 QY 67 GlyHisGlyThrHisValAlaGlySerValIleGlyValGlyValGlyValGlyValGly 84
 Db 595 GGACACGGACCCACGTTGGGGTATCGTTGCCGGAACCGGACGGTTAACTCCACGTAC 654

QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
 Db 655 ATAGCGTCTCGCCCGCGCGGAGAGTCTCGCGGTCAAGTTCTCGTCCGCGAGGTTCG 714
 QY 102 GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrrSerAla 121
 Db 715 GGAAGCGTCTCCACCATCATCGCGGTGTTGACTCGGTCTCGTCCAGAACAGGACAGTAC 774
 QY 122 GlyAlaArgIle-----HisThrAsnSer 129
 Db 775 GGGATAAGGGTCAACAACCTCTCCCTCGGTCTCTCCAGAGTCCGACGGAACCGACTCC 834
 QY 130 TrpGlyAlaAlaValAsnGlyAlaTyrrThrThrAspSerArgAsnValaAspTyrrVal 149
 Db 835 CTCAGTCAGGCGGTCAACAACGCTGGGACGCC----- 867
 QY 150 ArgIysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr 169
 Db 868 -----GGTATAGTAGTCTCGTCTCGCGCGGCAACAGCGGCGGCAACACTACACC 918
 QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
 Db 919 GTCGGCTCACCGCGCGCGGAGCAAGGTCAACCGTCGTGCA----- 963
 QY 190 ProSerPheGlySerTyrrAlaAspAsnIleAsnHisValAlaGlnPheSerArgGly 209
 Db 964 -----GTTGACAGCAGCAGCAACATCCAGCTCTCCAGCAGGGA 1005
 QY 210 ProThrIysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
 Db 1006 CCGACCCGCGGAGCGGTCAAGCGGAAGTCTGTCGCCCGCGGTGTCATCATAGCC 1065
 QY 230 AlaArgSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerIleTyrrAla 249
 Db 1066 CCGCGCGCCAGC-----GGAACGAGCATGGGACCCCGGATAACAGCTACTACACC 1116
 QY 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu 268
 Db 1117 AAGCCTCTGGACCAACATGGCCACCCCGCAGTTTCGGGCGCTTGGCGGCTCATCTC 1176
 QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
 Db 1177 CAGGCCAC-----CCGAGCTGGACCCCGGCAAGGTGAAG 1212
 QY 287 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
 Db 1213 ACCGCTCATCGAGACCGCGGACATAGTCGCCCGGAGAGATAGCGGACATCGCCTAC 1272
 QY 299 GlyTyrrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- 316
 Db 1273 GGTGG-----GGTAGGGTGAACGCTCTACAAGGCGCATCAAGTAC 1311
 QY 317 ---AsnValAlaTyrrValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrr 335
 Db 1312 GACGACTACGCCAAGCTACCTTCACCGGTCTCGCTCGCCGACAGGGAAGCGGCCAC 1371
 QY 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValThrPheAspAlaPro 355
 Db 1372 ACCTTCGACGCGCGCGGCCACCTTCGTGACCGCCACCTCTACTCTGGAC----- 1422
 QY 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
 Db 1423 -----ACGGGTCTGAGGACATCGACCTCTACCTCTACGACCCC 1461
 QY 376 AsnGlyThrGlnTyrrValGlyAsnAspPheThrSerProTyrrAsnAspAsnTrpAspGly 395
 Db 1462 AACGGGAACGAG---GTTGACTACTCTTACACCGCTACTAC----- 1500
 QY 396 ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrrThrIleGlu 415
 Db 1501 -----GGCTTCGAGAGGTCTGCTACTACAAACCCCGGACCTCGACGCTCAAG 1554

QY 416 ValcHnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 DB 1555 GTGCTCAGCTACCAAG-----GGCGCGGCAACTACCAAGCTCAGCTCGTCAGC 1602

RESULT 8

AA05920
 ID AAX05920 standard; DNA; 1236 BP.

XX AAX05920;

XX AC

XX 06-MAY-1999 (first entry)

XX Hyperthermostable protease fragment encoding DNA.

XX Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.

XX Pyrococcus furiosus.

XX WO9856926-A1.

XX 17-DEC-1998.

XX 04-JUN-1998; 98WO-JP002465.

XX 10-JUN-1997; 97JP-00151969.

XX (TAKI) TAKARA SHUZO CO LTD.

XX Takakura H, Morishita M, Shimojo T, Asada K, Kato I;

XX WPI; 1999-080907/07.

XX P-PSDB; AA094836.

XX Recombinant hyperthermostable protease from Pyrococcus furiosus - and
 PT gene encoding it, for large scale production of the protease for
 PT industrial use.

XX Claim 6; Page 37-38; 82pp; Japanese.

XX The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
 CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially bacillus strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis

SQ Sequence 1236 BP; 368 A; 251 C; 309 G; 308 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 5.85e-23 Length: 1236
 Score: 416.50 Matches: 141
 Percent Similarity: 42.55% Conservative: 59
 Best Local Similarity: 30.00% Mismatches: 148
 Query Match: 18.54% Indels: 122
 DB: 2 Gaps: 20

US-09-985-689A-1 (1-434) x AAX05920 (1-1236)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24

DB 24 GTCTGCAGCTCAAGTATATGGCACTTACGTTTGGAACTTGGATATGATGTTCTGGAAT 83

QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44

DB 84 CACAATAGGAATAATATGACACTGGAATTGAC-----GCTTCTCATCC 125

QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 DB 126 AGATCTCCAGGAAAGTA-----ATTGGGTGGGTAGATTTTGTCAATGG 170
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
 DB 171 TAGGAGTTATCCATACGATGACCATGGACATGGAACTCATGTAGCTTCAATAGCAGCTGG 230
 QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 DB 231 TACTGGAGCAGCAAGTAATGCAAGTACAGGGAAATGGCTCCAGGAGTAAGCTGGGGG 290
 QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
 DB 291 AATTAAAGTTCTAGTGCCGATGTTCTTGAAGCATATCTACTATATAATTAAGGAGTTGA 350
 QY 112 mThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
 DB 351 GTGGGCCCTTGATACAAAGATAAGTACGGAATTAAAGTCAATTAACTCTTCTCTGGTTC 410
 QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
 DB 411 AAGCCAGAGCTCAGATGGTACTAGCTCTAAGTCAGCTCTTAAATGCAGCGTGGATGC 470
 QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
 DB 471 T-----GGATTAGTTGTTGGTTCGCCG 494
 QY 160 eGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 180
 DB 495 TGGAAACAGTGGACCTTAACAGTATACAATGGTTCTCCAGCAGCTGCAGCAAGTTAT 554
 QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
 DB 555 TACAGTTGGAGCC-----GTTGACACAGTATGA 581
 QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 DB 582 TGTATAACAAGCTTCTCAAGCAGAGGGCCAACTGCAGCGCAGCTTAAGCCCTGAGGT 641
 QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
 DB 642 TGTGCTCCAGAAACTGGAATAATTGCTGCCAGACAAAGT-----GGAACATGACAT 692
 QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProI1 260
 DB 693 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTCTGGACATCAATGGCAACTCTCA 752
 QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
 DB 753 CGTAGCTGGTATGAGCGCCCTCTGCTCAA-----GCACACCC 791
 QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
 DB 792 GAGCTGGACTCCAGACAAAAGTAAACACAGCCCTCATAGAAACTGCTGATATCGTAAAGCC 851
 QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
 DB 852 AGATGAATAGCCGATATAGCCCTACGGTGCA-----GGTAGGGT 890
 QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
 DB 891 TAATGCATACAAGGCTATAAAC-----TACGATAACTATGCAAGAGTAGTGTTCACCTGG 944
 QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
 DB 945 ATATGTTGCCAACAAAGCGCAGCAAACTCCAGGTTCTGTTATAGCGAGCTTCGTTGCT 1004
 QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 DB 1005 AACTGCCACATATATACTGGACAATGCCAAT----- 1035
 QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385

[illegible]

Db 891 TAATGCATACAGGCTATAAAC-----TACGATAACTATGCAAGCTAGTGTCTACTGG 944
Qy 330 rGlnLyAala-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
Db 945 ATATGTTGCCAACAAGCAGCACAACCTCACCAGTTCGTTATTAGCGGAGCTTCGTTGCT 1004
Qy 345 uLysLeSerLeuValTrpSerPheAlaProAlaSerThrAlaSerValThrLeuVa 365
Db 1005 AATGCCACATTAATACGGGACATGCAAT----- 1035
Qy 365 lAsnAspLeuAspLeuValThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
Db 1036 -AGGACCTTGATCTTTACCTCTACGATCCCAATGGAAACAG---GTTGACTACTCTTA 1091
Qy 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheLeAs 405
Db 1092 CACGCCCTACTAT-----GGATTCCAAAAGGTTGGTTATTA 1127
Qy 405 nAlaProGlnSerGlyThrTyrThrThrLeuGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAACCAACTGATGGACATGGACAAATTAAGTTGTAGCTACAGC-----GGAAG 1178
Qy 425 oGlnThrPheSerLeuAlaLeValAsn 434
Db 1179 TGCAAACTATCAAGTAGATGTGGTAAGT 1206
RESULT 10
ID AAT85695 standard; DNA; 1962 BP.
AC AAT85695;
XX 17-OCT-2003 (revised)
DT 20-APR-1998 (first entry)
XX Pyrococcus furiosus PFUS protease coding sequence.
DE Pyrococcus furiosus PFUS protease coding sequence.
XX Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
XX Pyrococcus furiosus; DSM-3638.
XX WO9721823-A1.
XX 19-JUN-1997.
XX 07-NOV-1996; 96WO-JP003253.
XX 12-DEC-1995; 95JP-00323285.
XX (TAKI) TAKARA SHUZO CO LTD.
XX Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
PI Tsunawawa S, Kato I;
XX WPI: 1997-332794/30.
XX P-PSDS; AAW24129.
XX Protease(s) and genes encoding them obtained from Thermococcus and
PT Pyrococcus strains - have extremely high thermal stability and are useful
PT industrially and as research reagents.
XX Disclosure; Page 123-125; 159pp; Japanese.
XX This sequence represents the coding sequence for the protease from
CC Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
CC invention. The proteases of the invention have extremely high thermal
CC stability. The proteases can be used as research reagents, and
CC industrially in the food, drug and chemical industries. (Updated on 17-
CC OCT-2003 to standardise OS field)
XX Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.01e-22 Length: 1962
Score: 416.50 Matches: 141
Percent Similarity: 42.55% Conservative: 59
Best Local Similarity: 30.00% Mismatches: 148
Query Match: 18.54% Indels: 122
DB: 2 Gaps: 20
US-09-985-689A-1 (1-434) x AAT85695 (1-1962)
Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGly 24
Db 420 GTCTGCAGCTCAAGTATTATGGCAACTTACGTTGGAACTGGGATATGATGTTCTGGAA 479
Qy 24 nLeValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG 44
Db 480 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysLeuThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAGAAAGTA-----ATTGGTGGGTAGATTTTGTCAATGG 566
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG 78
Db 567 TAGGAGTTATCCATACGATGACCATGGACATGGAACATCATGTAGCTTCAATAGCAGCTG 626
Qy 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGCAAGTACAGGGAATGCTCCAGAGGCTAAGCTGGCGG 686
Qy 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuG 112
Db 687 AATTAAGGTTCTAGTGCCGATGTTCTGGAAGCATATCTACTATAATTAAGGAGGTGA 746
Qy 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGCCCTTCATACAAAGATAAGTACGGAATTAAGTCAATATCTTCTCTGTGTC 806
Qy 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 807 AAGCCAGAGCTCAGATGGTACTGACGCTTAAGTCAGCTGTTAATGCAGCGTGGGATGC 866
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 T-----GGATTAGTTGTTGTTGCTGCGC 890
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI 180
Db 891 TGGAAACAGTGGACTTAAAGTATACATCGTTCTCCAGCAGCTGCAAGCAAGTTAT 950
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 951 TACAGTTGGAGCC-----GTTGACAAAGTATGA 977
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 978 TGTATAAACAAGCTTCTCAAGCAGAGGCGCAACTGCGAGCGGAGGCTTAAGCTGAGGT 1037
Qy 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTGTCTCCAGAAACTGGATAAATTGCTGCCAGACGAAGT-----GGAACATGACAT 1088
Qy 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIl 260
Db 1089 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTCTGGGACATCAATGGCACTCTCA 1148
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 1149 CGTAGCTGTATTGACGCCCTCTTGTCTCCA-----GCACACCC 1187
Qy 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 1188 GAGCTGGACTCCAGACAAAGTAAAGAACAGCCCTCATAGAAACTGCTGATATCGTAAAGCC 1247

QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyValGva 310
 Db 1248 AGATGAATAGCGATATAGCTACGGTCCA-----GGTAGGTT 1286
 QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
 Db 1287 TAATGCATCAAGGCTATAAAC-----TACGATAACTATGCAAGTAGTGTTCACCTGG 1340
 QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
 Db 1341 ATATGTTGCGCAACAGGCGAGCAACTCACCGATCGTTATTACGGGAGCTTCGTTGTT 1400
 QY 345 uLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 Db 1401 AACTGCCACATTATCTGGGCAATGCCAAT-----TACGATAACTATGCAAGTAGTGTTCACCTGG 1431
 QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
 Db 1432 -AGCGACCTTGATCTTACTCTACGATCCCAATGGAACCCAG---GTGACTACTCTTA 1487
 QY 385 eThrSerProTyrAsnAspAsnTyrAspGlyArgAsnAsnValGluAsnValPheIleAs 405
 Db 1488 CACCGCCTACTAT-----GGATTGGAAGGTTGGTTATTA 1523
 QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
 Db 1524 CACCCCACTGATGGACATGACATTAAGTTGTAGCTACAC-----GGAAG 1574
 QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
 Db 1575 TGCAAACTATCAAGTAGATGTGTAAGT 1602

RESULT 11

AA05929
 ID AA05929 standard; DNA; 1962 BP.

AC AAX05929;
 XX
 DT 06-MAY-1999 (first entry)
 XX

DE Hyperthermostable protease encoding DNA.

XX Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.

XX Pyrococcus furiosus.

XX WO9856926-A1.

XX 17-DEC-1998.

XX 04-JUN-1998; 98WO-JP002465.

XX 10-JUN-1997; 97JP-00151969.

XX (TAKI) TAKARA SHUZO CO LTD.

XX Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
 XX WPI; 1999-080907/07.

XX P-PSDB; AAW94841.

PT Recombinant hyperthermostable protease from Pyrococcus furiosus - and
 PT gene encoding it, for large scale production of the protease for
 PT industrial use.

PS Disclosure; Page 59-60; 82pp; Japanese.

XX The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
 CC working temperature 40-110 deg.C (Optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at

CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially Bacillus strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis

XX SQ Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,01e-22 Length: 1962
 Score: 416.50 Matches: 141
 Percent Similarity: 42.55% Conservative: 59
 Best Local Similarity: 30.00% Mismatches: 148
 Query Match: 18.54% Indels: 122
 DB: 2 Gaps: 20

US-09-985-689A-1 (1-434) x AAX05929 (1-1962)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGly 24
 Db 420 GTCTGCAGCTCAAGTTATGCAACTTACGTTTGGAACTTGGGATATGATGTTCTGGAAT 479
 QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG 44
 Db 480 CACAATAGGATATATGACACTGGAATTGAC-----GCTTCTCATCC 521
 QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 Db 522 AGATCTCCAAGGAAAGTA-----ATTGGTGGGTAGATTTCATCAATGG 566
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG 78
 Db 567 TAGGAGTTATCCATACATGACCATCGACATGGAATCATGTAGCTTCATAGCAGCTGG 626
 QY 78 YAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 Db 627 TACTGGAGCAGCAAGTAATGGCAAGTACAAGGGAATGGCTCCAGGAGCTAAGCTCGGG 686
 QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG 112
 Db 687 AATTAAGTTCTAGGTCGCGATGTTCTGGAAGCATATCTACTATATTAATTAAGGAGTTGA 746
 QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
 Db 747 GTGGCGCGTTGATAACAAAGATAAGTAGCGAATTAAGGTCATTAACTCTTCTTGGTTC 806
 QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
 Db 807 AAGCCAGAGCTCAGATGGTACTGACGCTTAAGTCAGGCTGTTAATGCGAGCTGGATGC 866
 QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
 Db 867 T-----GGATTAGTTGTTGTTGCTGCGC 890
 QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
 Db 891 TGGAAACAGTGGACCTAACAAAGTATACAATCGGTTCTCCAGCAGCTGCAAGCAAGTTAT 950
 QY 180 eThrValGlyValaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
 Db 951 TACGTTGGAGCC-----GTTGACCAAGTATGA 977
 QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 Db 978 TGTATAACAAGCTTCTCAAGCAGAGGGCCAACTGCAGACGGCAGGCTTAAGCTGAGGT 1037
 QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPh 240
 Db 1038 TGTTCTCCAGGAACCTGGATTAATGCTGCAGAGCAAGT-----GGAAGTACAT 1088

QY 240 eTTPAlaAsnHisaspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIl 260
 Db 1089 GGGTCAACCAATTATGACTATTACACAGCAGCTCTGGGACATCAATGGCACTCTCA 1148
 QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeThrPr 280
 Db 1149 CGTAGCTGGTATTGCGCCCTCTTGCTCCAA-----GCACACCC 1187
 QY 280 oLys-----ProSerLeuLysAlaAlaLeuIleAlaGlyAla----- 293
 Db 1188 GAGCTGGACTCCACAAAGTAAAAAGCCCTCATAGAACTCTGATATCGTAAGCC 1247
 QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgVa 310
 Db 1248 AGATGAATACCGATATAGCTACGGTGCA-----GGTAGGGT 1286
 QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
 Db 1287 TAATGATACAGAGCTATPAAAC-----TAGATAACTATGCAAGCTAGTGTCTACGG 1340
 QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
 Db 1341 ATATGTTGCCAACAAAGGCGCAAACTCACCAGTTCGTTATTAGCGGAGCTTCGTTCTGT 1400
 QY 345 uLysIleSerLeuValTTPSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 Db 1401 AACTGCCACATATACTGGGCAATGCCAAT----- 1431
 QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
 Db 1432 -AGCGACCTTGATCTTTACCTCTAGACTCCCAATGGAAACCAG--GTTGACTACTCTTA 1487
 QY 385 eThrSerProTyrAsnAspAsnTTPAspGlyArgAsnAsnValGluAsnValPheIleAs 405
 Db 1488 CACCGCTTACTAT-----GGATTCGAAAGGTGGTTATTA 1523
 QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
 Db 1524 CAACCCACTGATGGAACTGGACATTAAGTTGTAAAGCTACAGC-----GGAAG 1574
 QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
 Db 1575 TGCAAACTATCAAGTAGATGTGGTAAAGT 1602

RESULT 12

AAT85669

ID AAT85669 standard; DNA; 1977 BP.

XX AC AAT85669;

XX DT 20-APR-1998 (first entry)

XX DE Protease coding sequence.

XX KW Protease; research reagent; thermal stability; ss.

XX OS Synthetic.

XX PN WO9721823-A1.

XX PD 19-JUN-1997.

XX PF 07-NOV-1996; 96WO-JP003253.

XX PR 12-DEC-1995; 95JP-00323285.

XX PA (TAKI) TAKARA SHUZO CO LTD.

XX PI Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;

XX PI Tsunawasa S, Kato I;

XX DR WPI; 1997-332794/30.

DR P-PSDB; AAW24123.

XX

Protease(s) and genes encoding them obtained from Thermococcus and
 Pyrococcus strains - have extremely high thermal stability and are useful
 industrially and as research reagents.

PS Claim 11; Page 95-97; 159pp; Japanese.

XX This sequence represents the coding sequence for a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries

XX SQ Sequence 1977 BP; 527 A; 562 C; 513 G; 375 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1,02e-21	Length:	1977
Score:	403.50	Matches:	134
Percent Similarity:	43.07%	Conservative:	68
Best Local Similarity:	28.57%	Mismatches:	147
Query Match:	17.96%	Indels:	120
DB:	2	Gaps:	20

US-09-985-689A-1 (1-434) x AAT85669 (1-1977)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyCl 24
 Db 420 GTCTGCGAGCTCAAGTATATGGCACTTACGTTTGGAACTTGGGATATGATGGTCTGGAAAT 479

QY 24 nileValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGl 44
 Db 480 CACAATAGGAATAATTGACACTGGAATTGAC-----GCCTCTCATCC 521

QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 Db 522 AGATCTCCAAGGAAAGTA-----ATTGGTGGGTAGATTTTGTCAATGG 566

QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValleuGl 78
 Db 567 TAGGAGTTATCCATCAGTACCATGGACATGGACATGACTCATGTAGCTTCAATAGCAGCTGG 626

QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 Db 627 TACTGGAGCAGCAAGTAATAGCAAGTACAAAGGAATGGTCCAGGAGCTAAGCTGCGCGG 686

QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyGlyLeuProSerAsnLeuGl 112
 Db 687 AATTAAGTTCTAGGTGCGGATGGTCTGGAAGCAATCTACTATAATAAGGAGGTGA 746

QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgile----- 125
 Db 747 GTGGGCGGTGATACAAAGATAAGTACGGAATTAAAGTCAATATCTTCTTGGTTC 806

QY 126 -----HisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrTh 140
 Db 807 AAGCCAGAGCTCCGAGCAACCGACTCCCTCAGTCCGCGGTCAACACGCTCGGAGCGC 866

QY 140 rAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
 Db 867 C-----GGTATAGTAGTCTGGTCCGCGC 890

QY 160 aglyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
 Db 891 CGGCAACAGCGGCGGCAACACCTACACCGTCCGCTCACCCTCCGCGGCGGAGCAAGGTGAT 950

QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
 Db 951 AACCGTCGGTGCA-----GTTGACAGCAACGA 977

QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 Db 978 CAACATCGCCAGCTTCTCCAGCAGGAGCCGACCCGCGGAGCGGAGGCTCAAGCCGGAAGT 1037

QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240

```

1038 CGTCGCCCGCGGCTTGACATCATAGCCCGCGCCAGC-----GGAAACAGCAT 1088
1039
240 eTPAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProI 260
1089 GGGCACCAGATTAACAGTACTACTACCAAGGCTCTGGAACAGCATGGCCACCCGCA 1148
260 eValAlaGly---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeu 279
1149 CGTTTCGGCGGCTTGCGCGCTCATCTCCAGGCCAC----- 1185
279 rProLys-----ProSerLeuLeuLysAlaAlaLeuLysAlaGlyAla----- 293
1186 -CGAGCTGAGCCCGGACAGGTGAAGACCGCCCTCATCGAGACCGCCGACATAGTCG 1244
294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyAr 309
1245 CCCCAAGGAGATAGCGGACATCGCTACGGTGG-----GGTAG 1283
309 gValThrLeuAspLysSerLeu-----AsnValAlaTyrValAsnGluSerSe 326
1284 GTGAACGCTCTACAGGCCATCAAGTACGACGACTACGCCAAGCTCACCCTCACCGGCTC 1343
326 rLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThrAlaGlyLysProLeu 346
1344 CGTCGCCGACAGGGAAGCGCCACCCACCTTCGACGTCAGCGCGCCACCTTCGTGAC 1403
346 sIleSerLeuValTyrSerAspAlaProLaserThrThrAlaSerValThrLeuVal 366
1404 CGCACCCCTTACTGGAC-----ACGGGCTCGAG 1433
366 nAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPhe 386
1434 CGACATCGACCTTACTCTACGCCCCACCGGAACGAG-----GTTGACTACTCCTAC 1490
386 rSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsn 406
1491 CGCTACTAC-----GGCTTCGAGAAGTCGGTACTACAA 1526
406 aProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPro 426
1527 CCGACCGCGGAGACCTGGACGTCAGGTCGACGTACAG-----GGCGCGC 1577
426 nThrPheSerLeuAlaIleValAsn 434
1578 GAATACACAGTCTGACGCTGTCAGC 1602

RESULT 13
ABL54900
ID ABL54900 standard; DNA; 2121 BP.
AC ABL54900;
XX
XX
XX
XX
11-SEP-2003 (revised)
DT
31-MAY-2002 (first entry)
DE
T. yonsei subtilisin-like serine protease coding sequence.
XX
Subtilisin-like serine protease; ss.
KW
XX
XX
Thermoanaerobacter yonseiensis.
XX
XX
Key Location/Qualifiers
FT CDS
FT 142..179
FT /*tag= a
FT /product= "subtilisin-like serine protease"
XX
XX
KR2000072141-A.
XX
XX
05-DEC-2000.
PD
XX
XX
04-AUG-2000; 2000KR-00045411.
PF
XX

```

```

PR 04-AUG-2000; 2000KR-00045411.
XX
XX (KIMY/) KIM Y S.
XX
XX Chang HJ, Kim DH, Byun YR, Kim YS;
XX
XX WPI; 2001-298092/31.
XX
XX P-ESDB; ABB09483.
XX
XX
XX New DNA sequence of thermophilic protein decomposition enzyme and protein
XX derived therefrom.
XX
XX Claim 1; Page 6; 15pp; Korean.
XX
XX This sequence represents the DNA encoding the Thermoanaerobacter yonsei
XX subtilisin-like serine protease of the invention. (Updated on 11-SEP-2003
XX to standardise OS field)
XX
XX Sequence 2121 BP; 712 A; 410 C; 425 G; 554 T; 0 U; 20 Other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 2.96e-19 Length: 2121
XX Score: 372.00 Matches: 130
XX Percent Similarity: 46.29% Conservative: 51
XX Best Local Similarity: 33.25% Mismatches: 114
XX Query Match: 16.56% Indels: 96
XX DB: 4 Gaps: 19
XX
XX US-09-985-689A-1 (1-434) x ABL54900 (1-2121)
XX
XX QY 6 GlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGly 25
XX Db 445 GGAATCAAAAA-----GCACGGAGTGATTTTGGAGTCACAGGAAAAATATAACA 495
XX QY 26 ValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAla 45
XX Db 496 ATAGCAATTATGACAGGTATAGCGAAATACAGTTGACCTCTCA----- 543
XX QY 46 PheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn----- 63
XX Db 544 --CGTGAAAAATA-----ATAGGATGGAAGACATTATCAACAA 585
XX QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
XX Db 586 ACTACACATACGACGACAAATGCCATGGAACCTCAGTACGAACTATTCGTGAGGTACA 645
XX QY 80 GlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSer 96
XX Db 646 GGTCTGGAACACAGCTCTTACAAAGCGGTGCTCCTGATGCTTTGTTGGTGAATAAAA 705
XX QY 97 IleMetAspSerGlyGly-----LeuGlyGlyLeuProSerAsn 110
XX Db 706 GTTTTAGATGCAATGGAAGCGGACGATGAGCAGCTGTAACCTGCGGAATTGACTGGGCT 765
XX QY 111 LeuGln-----ThrLeuPheSerGlnAla 118
XX Db 766 GTTCAAAATAAGATGTATACGGAAATCAAGTTATAAATTAGCTTCGCACTTCTACA 825
XX QY 119 TyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyr 138
XX Db 826 AGTTCTGATGGA-----ACTGACTCTACCTCATTAGCAGTGAAT----- 864
XX QY 139 ThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeuPhe 158
XX Db 865 -----AGGCAGTAGAT-----ACCGGTATTGTAGTAGTTGTA 897
XX QY 159 AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsn 178
XX Db 898 GCACGAGAAACTCTGGCCCTGCAAAATACACCATAGGTCCTCCCTGTGTGCGGAAAAA 957
XX QY 179 AlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerThrAlaAspAsn 198
XX Db 958 GCCATAACAGTCGACGAATGGCAGATGTA-----GGTGAACCTTGGCTTTAAC 1005

```

QY 199 IleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysPro 218
 Db 1006 CTT-----GCAGCTTTTCAGCGCGGTCTACTCTGCTGACGGAGAAATAAACCT 1056
 QY 219 AspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer 238
 Db 1057 GACATTGGCGCCAGGATATAATAAATGCTGCCCAAG----- 1095
 QY 239 SerPheTrpAlaAsnHisAspSerSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThr 258
 Db 1096 -----GCAATTCGTGAATGGATATGTAACATACAGCGGTACACGATGGCAACA 1146
 QY 259 ProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIle 278
 Db 1147 CTTTGTAGCAGAACTGTTCTTATGCTTAAC-----GCTAATCCAAATCTC 1197
 QY 279 ThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeu 298
 Db 1198 ACTCCAATGATGCA-----AAAATATAATATGTTACTGCAAAAAGCTGGGGCCT 1251
 QY 299 GlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnVal 318
 Db 1252 CCAAGCAAAACGTTACTATGTCGAGGAAGATGGCTGATGAAGCTATAAGGGTA 1311
 QY 319 Ala-----TyrValAsnGluSer 324
 Db 1312 GCAGGTAATTTAGAGAAATAATATTGATGTACCAAAATCATTATTATATT-----TCA 1365
 QY 325 SerSerLeuSerThrSerGlnLysAla---ThrTyrSerPheThrAlaThr---AlaGly 342
 Db 1366 GGTATCTGCCGGTTCTCGTACAGTGATCTGGACTTTTAACGCAACATAACAAGT 1425
 QY 343 LysProLeuLysIleSerLeuValTrpSerAsp 353
 Db 1426 TATCCAATCGCAATAACATTAAATAATTCCTGAC 1458
 RESULT 14
 AAT61454
 ID AAT61454 standard; DNA; 2539 BP.
 XX AC AAT61454;
 XX DT 06-OCT-1997 (first entry)
 XX DE Streptomyces viridosporus dhpa gene.
 XX KW asymmetric hydrolase, dhpa; 4-substituted-1,4-dihydropyridine;
 KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
 KW cardiovascular; treatment; hypertension; ischaemic heart disease; ds.
 XX OS Streptomyces viridosporus.
 FH Key Location/Qualifiers
 FT CDS 338..2539
 FT /tag= a
 FT /note= "no stop codon given"
 FT misc_feature 950..2509
 FT /tag= b
 FT /note= "encodes AAW13666"
 XX WO9705243-A1.
 XX 13-FEB-1997.
 XX 30-JUL-1996; 96WO-JP002147.
 XX 31-JUL-1995; 95JP-00212975.
 XX 29-FEB-1996; 96JP-00067478.
 XX (SAOC) MERCIAN CORP.
 XX Arisawa A, Matsufuji M, Tsuruta T, Dobashi K, Nakashima T;
 PI

PI Isshiki K, Yoshioka T;
 XX WPI: 1997-145582/13.
 DR P-PSDB; AAW13666, AAW13667.
 XX
 PT Asymmetric hydrolase gene derived from Streptomyces viridosporus - acts
 on 4-substituted-1,4-dihydropyridine derivatives to produce chiral
 derivatives useful for synthesis of cardiovascular drugs.
 XX Claim 3; Page 49-55; 78pp; Japanese.
 CC This sequence is the Streptomyces viridosporus dhpa gene which encodes an
 asymmetric hydrolase which acts on 4-substituted-1,4-dihydropyridine
 derivatives. The enzyme allows the efficient conversion of 4-substituted-
 1,4-dihydropyridine esters to chiral partially hydrolysed derivatives,
 for use in the synthesis of cardiovascular drugs suitable for the
 treatment of e.g. hypertension and ischaemic heart disease
 XX Sequence 2539 BP; 433 A; 959 C; 867 G; 280 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3,686-17 Length: 2539
 Score: 346.00 Matches: 129
 Percent Similarity: 42.00% Conservative: 52
 Best Local Similarity: 29.93% Mismatches: 167
 Query Match: 15.40% Indels: 84
 DB: 2 Gaps: 13
 US-09-985-689A-1 (1-434) x AAT61454 (1-2539)
 QY 2 AspValAlaArgGlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
 Db 953 GACAGCTGCTCGGCGAGATCGCGCCCCCAAGGCTGTGTCGCC---GGTACGACGCG 1009
 QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
 Db 1010 AAGGCGGTGAAGATCGCGTCTCTGGACACCGGTGTGACACGAGC----- 1054
 QY 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
 Db 1055 ---CATCGGACCTGAAGGCGCGGTGACCGCGTCCAAAGACTTCACCGCGCGCGCGCC 1111
 QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
 Db 1112 GCGCGCGACAAGGTGGCGCCACGCCACCCACGTCGCTCGATCGCGCGCGCGCGCGCC 1171
 QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
 Db 1172 CAGTCCAAAGGCGCAAGTACAAGGCGGTGCGACCGCGCGCGGATCCTCAACGCGCAAGTC 1231
 QY 98 MetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
 Db 1232 CTCGACGACTCGGT-----TTCGGCGACGACTCCGCGATCCTCGCGCGATGGAGTGG 1285
 QY 118 AlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
 Db 1286 GCGCGCGCGCGAGCGCGCGAGCTGCTCAACATGAGCTCGCGCGCGATCGACACACCGGAG 1345
 QY 137 aTyrThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLe 157
 Db 1346 ACCGACCGCTGGAGCGCGCG-GTCCGACAAGCTGCCCGCGAGAGGCGTCTGTTCGC 1404
 QY 157 uPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLy 177
 Db 1405 CATCGCGCGCGCAACGAGGCGCGGAG-----TCGATCGTTCGCGCGCGCGCGGA 1458
 QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
 Db 1459 CGCCGCGCTTCAACGCGCGCGC-----GTGCA 1485
 QY 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgIle 216
 Db 1486 CGACAAGGACCAAGCTCGCGACTTCTCTCCACCGCGCGCGCGCGCGCAT 1545


```

QY 216 eLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr 236
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1546 CAAGCCGGAGCTCACCGCTCCCGCGTGGACATCACGGCGCTCGGCGAGGCAACGA 1605
QY 236 oAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMe 256
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1606 CATCGGCCAGGAGTCCGTGAGGACCGCGCGCTACATGACCATCTCGGCGACGTCGAT 1665
QY 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1666 GCGACCCCGACGTCGCGCGCGCGCGCTCTCTGAAGCAGCAG----- 1711
QY 276 gGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl 293
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1712 -----CACCCCGACTGGACCTCCGCGCAACTGAAGGGCGCGCTCACCGGCTC 1758
QY 293 aAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAs 313
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1759 CACCAAG---GGCGGCAAGTACACCCCGTTCGAGCAGGTTTCGGCGCGGATCCAGGCCGA 1815
QY 313 pLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSer----- 328
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1816 CAAGGCGCTCCACGACGATGATCGCGACCGCGTCTCGTGAGCTTCGGCGTCCAGCA 1875
QY 329 -----ThrSerGlnLysAlaThrTyrSerPheThrAlaTh 340
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1876 GTGGCGGCACACGACGACGAGCGGTCCACCAAGCAGCTGACCTACCGCAACTCGGCAC 1935
QY 340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1936 CCAGGAGCTCACGCTGAGCTGACGTCGACCGCCACCGACCCCAAGGGCAAGCGGCCCC 1995
QY 354 -----AlaProAlaSerThrThrAlaSe 361
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1996 GCGGCGCTCTTCACGCTCGCGGCCACACCGGTGACCGTCCCGCGCGCGCGCGCTC 2055
QY 361 rValThrLeuValAsnAspLeuAspLeu-----Va 371
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2056 CGTCGACATGACCGCGCACACCGCGTTCGGCGGCGGCGGTGGACGGCGGTACTCGCGCTA 2115
QY 371 lIleThrAlaProAsnGlyThrGlnTyrVal 381
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2116 CGTGTGCGCACCGCGCGCGCGCGCGTCC 2146

RESULT 15
AAr61455
ID AAr61455 standard; DNA; 2809 BP.
XX AC AAr61455;
XX AC AAr61455;
DT 17-OCT-2003 (revised)
DT 06-OCT-1997 (first entry)
XX DE DhpA-mel chimeric gene.
XX XX
KW asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;
KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
KW cardiovascular; treatment; hypertension; ischaemic heart disease; ds.
XX OS Streptomyces viridosporus.
OS Streptomyces antibioticus.
OS Chimeric.
XX XX
FH Key Location/Qualifiers
FH CDS 338..2809
FT /*tag= a
FT mat_protein 338..2539
FT /*tag= b
FT /*product= "DhpA_protein_product"
FT /*notes= "from S. viridosporus"
FT mat_protein 2540..2809
FT /*tag= c

```

```

FT /product= "Melanin"
FT /note= "from S. antibioticus"
XX W09705243-A1.
XX 13-FEB-1997.
XX 30-JUL-1996; 96WO-JP0021147.
XX 31-JUL-1995; 95JP-00212975.
XX 29-FEB-1996; 96JP-00067478.
XX (SAOC) MERCIAN CORP.
XX Arisawa A, Matsufuji M, Tsuruta T, Dobashi K, Nakashima T;
XX Isshiki K, Yoshioka T;
XX MPI; 1997-145682/13.
XX P-FSDS; AAW13665.
XX Asymmetric hydrolase gene derived from Streptomyces viridosporus - acts
XX on 4-substituted-1,4-di:hydro:pyridine derivatives to produce chiral
XX derivatives useful for synthesis of cardiovascular drugs.
XX Claim 5; Page 37-43; 78pp; Japanese.
XX This sequence is a fusion gene encoding Streptomyces viridosporus dhpA
XX gene product (an asymmetric hydrolase) which acts on 4-substituted-1,4-
XX dihydropyridine derivatives, and melanin from S. antibioticus. The dhpA
XX enzyme allows the efficient conversion of 4-substituted-1,4-
XX dihydropyridine esters to chiral partially hydrolysed derivatives, for
XX use in the synthesis of cardiovascular drugs suitable for the treatment
XX of e.g. hypertension and ischaemic heart disease. (Updated on 17-OCT-2003
XX to standardise OS field)
XX SQ Sequence 2809 BP; 473 A; 1041 C; 984 G; 311 T; 0 U; 0 Other;

Alignment Scores:
Pred No.: 4,14e-17 Length: 2809
Score: 346.00 Matches: 129
Percent Similarity: 42.00% Conservatives: 52
Best Local Similarity: 29.93% Mismatches: 167
Query Match: 15.40% Indels: 84
DB: 2 Gaps: 13

US-09-985-689A-1 (1-434) x AAT61455 (1-2809)
QY 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
DB 953 GACACGTCCGTCGGCGCAGATCGCGCCGCCCAAGCGTGTGTCGCC---GGCTACGACGGC 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
DB 1010 AAGGCGGTGAAGATCGCGCTCTGGACACCGGTGTGCACACGAGC----- 1054
QY 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
DB 1055 ---CATCGGACCTGAAGGGCGGTGACCGCGTCCCAAGAACTTCCCGCGCGCGCGC 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaLysSerValLeuGlyAsnGlySer 81
DB 1112 GCGGCGACAAAGTGGGCGCACGCCACCTCGCTCGATCGCGGCGGCGCGCGCGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
DB 1172 CAGTCCAAGGGCAAGTACAGGCGGTGCGACCGCGCGCGCGATCCTCAACGCGCAAGGTC 1231
QY 98 MetAspSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
DB 1232 CTGACGACCTCCGGT-----TTCGGCGACACTCCGGCATCTCCCGCGCGCATGGAGTGG 1285
QY 118 AlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAenGlyAl 137
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```
Db 1286 GCGGCGCGCGGCGCGCGGCGTCAACATGAGCTGGCGGCGATGACACACCGGAG 1345
Qy 137 aYrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleIe 157
Db 1346 ACCGACCCGCTGGAGCGCGG-GTCGCAAGCTGCCGAGAGGCGGCTCTGTTCCG 1404
Qy 157 uPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys 177
Db 1405 CATCGCGGCGCGCAACGAGGCGCGGAG-----TCGATCGGTTCGCGCGGCGAGCGCGA 1458
Qy 177 ssnAlaIleThrValGlyValAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
Db 1459 GCGCGGCTCTACCGTGGGCGC-----GTGGA 1485
Qy 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro-----ThrLysAspGlyArgIle 216
Db 1486 CGACAGGACAACTCGCGGCTCTCTCCACCGGCGCGCGCTCGGCGGAGCGGCGCCAT 1545
Qy 216 eLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr 236
Db 1546 CAAGCGGAGCTCACCGCTCCGCGCGTGACATCACGCGCGCTCGGCGGAGGCGCAACGA 1605
Qy 236 oAspSerSerPheTyrAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMe 256
Db 1606 CATCGCGGAGGCTGGTGAGGACCGCGCGGCTACATGACCATCTCCGCGACGTCGAT 1665
Qy 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
Db 1666 GCGGACCGCGCACGTCCGCGGCGCGGCGCTCTCTGAAGCAGCAG-----1711
Qy 276 gGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCGGACTGGACCTCCGCGGAACTGAAGGGCGCGCTCACGGCTC 1758
Qy 293 aAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAs 313
Db 1759 CACCAAG---GGCGGAAGTACACCCCGTTCCGAGCAGGTTCCGGCGCGGATCCAGGCGGA 1815
Qy 313 pLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSer-----328
Db 1816 CAAGGCGCTCCAGCAGACCGTATCCGCGCGACCGGCTCTCGGTGAGCTTCGGGTCACGCA 1875
Qy 329 -----ThrSerGlnLysAlaThrTyrSerPheThrAlaTh 340
Db 1876 GTGCGCGCACACCGACGAGCGGTCACCAAGCAGCTGACCTACCGCAACCTCGGCAC 1935
Qy 340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp-----353
Db 1936 CCAGGACGTACGCTGAAGCTGACGTCCGACCGCCACCCGACCCCAAGGGCAAGGCGGCCCC 1995
Qy 354 -----AlaProAlaSerThrThrAlaSe 361
Db 1996 GCGGGGTCTTCACGCTGGGCGCCACCGGTGACGCTCCGCGCGGCGGCGGCGGCGCTC 2055
Qy 361 rValThrLeuValAsnAspLeuAspLeu-----Va 371
Db 2056 CGTCGACATGACCGCGCGACACCGCGCTCGGCGGCGACGCTGGACGCGCGCTACTCGGCGTA 2115
Qy 371 lIleThrAlaProAsnGlyThrGlnTyrVal 381
Db 2116 CGTGTGCTCCAGCGGCGCGGCGGCGGCGTAC 2146
```

Search completed: March 15, 2004, 23:35:55
Job time : 424 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 23:21:18 ; Search time 84 Seconds
(without alignments)
2867.246 Million cell updates/sec

Title: US-09-985-689a-1

Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAVNPVGPQTFLAIVN 434

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US0985689/runat_10032004_112904_19757/app_query.fasta_1.583
-DB=issued Patents NA -QMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOFCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0985689@cgn1_1_44@runat_10032004_112904_19757 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2247	100.0	1923	4	US-09-509-814A-5 Sequence 5, Appli
2	2242	99.8	1923	4	US-09-509-814A-7 Sequence 7, Appli
3	2183	97.2	1920	4	US-09-509-814A-3 Sequence 3, Appli
4	2125.5	94.6	3003	2	US-08-873-479-41 Sequence 41, Appli
5	452.5	20.1	1977	3	US-08-894-818B-2 Sequence 2, Appli
6	452.5	20.1	1977	3	US-09-445-472-11 Sequence 11, Appli
7	416.5	18.5	1236	4	US-09-445-472-2 Sequence 2, Appli
8	416.5	18.5	1566	3	US-08-894-818B-4 Sequence 4, Appli
9	416.5	18.5	1962	3	US-08-894-818B-34 Sequence 34, Appli
10	416.5	18.5	1962	4	US-09-445-472-15 Sequence 15, Appli
11	403.5	18.0	1977	3	US-08-894-818B-6 Sequence 6, Appli
12	351	15.6	2539	3	US-09-000-016-3 Sequence 3, Appli

13	351	15.6	2539	4	US-09-514-340-3 Sequence 3, Appli
14	346	15.4	2809	3	US-09-000-016-1 Sequence 1, Appli
15	346	15.4	2809	4	US-09-514-340-1 Sequence 1, Appli
16	310.5	13.8	2532	1	US-07-671-376C-4 Sequence 4, Appli
17	307.5	13.7	1859	3	US-08-894-818B-15 Sequence 15, Appli
18	307	13.7	2835	1	US-08-750-532-2 Sequence 2, Appli
19	307	13.7	4765	1	US-08-750-532-8 Sequence 8, Appli
20	307	13.7	4765	3	US-08-894-818B-7 Sequence 7, Appli
21	307	13.7	4765	4	US-09-445-472-5 Sequence 5, Appli
22	296	13.2	898	1	US-08-750-532-7 Sequence 7, Appli
23	290.5	12.9	564	1	US-08-750-532-11 Sequence 11, Appli
24	290.5	12.9	564	3	US-08-894-818B-14 Sequence 14, Appli
25	276	12.3	1306	4	US-09-966-921A-1 Sequence 1, Appli
26	276	12.3	1330	4	US-09-966-921A-5 Sequence 5, Appli
27	256.5	11.4	1194	4	US-09-328-352-3407 Sequence 3407, Ap
28	256.5	11.4	10216	2	US-08-875-154-1 Sequence 1, Appli
29	252	11.2	1628	4	US-09-634-238-212 Sequence 212, App
30	251	11.2	840	1	US-08-434-255-7 Sequence 7, Appli
31	251	11.2	840	1	US-08-459-967-7 Sequence 7, Appli
32	251	11.2	840	1	US-08-460-327-7 Sequence 7, Appli
33	251	11.2	840	1	US-08-459-871-7 Sequence 7, Appli
34	251	11.2	840	3	US-09-024-532-1 Sequence 1, Appli
35	251	11.2	840	4	US-09-104-623A-1 Sequence 1, Appli
36	251	11.2	840	4	US-09-019-532-1 Sequence 1, Appli
37	251	11.2	840	4	US-09-417-359A-1 Sequence 1, Appli
38	251	11.2	840	4	US-08-705-185-1 Sequence 1, Appli
39	251	11.2	840	4	US-09-338-746-1 Sequence 1, Appli
40	251	11.2	1110	1	US-08-434-255-5 Sequence 5, Appli
41	251	11.2	1110	1	US-08-459-967-5 Sequence 5, Appli
42	251	11.2	1110	1	US-08-460-327-5 Sequence 5, Appli
43	251	11.2	1110	1	US-08-459-871-5 Sequence 5, Appli
44	251	11.2	1191	1	US-08-434-255-3 Sequence 3, Appli
45	251	11.2	1191	1	US-08-459-967-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-509-814A-5
; Sequence 5, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAKKI, KATSUHIISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-5

Alignment Scores:
Pred. No.: 1.14e-224 Length: 1923
Score: 2247.00 Matches: 434
Percent Similarity: 100.00% Conservative: 0

```
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-985-689A-1 (1-434) x US-09-509-814A-5 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 619 AATGATGTTTGGCGTGGGATTTGTAAGCGGATGTTGGCTCAGACGAGCTACGGGTTGTAT 678
Qy 21 GlyGlnGlyGlnIleValAlaAspThrGlyLeuAspThrGlyLeuAspThrGlyLeuAspThr 40
Db 679 GGCAAGGACAGATCGTACGGTGGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 738
Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 739 TCGATCATGAACCTTCGCGGGGAAATTAATGATATATATGATGCGGATGCGGATGCGGAT 798
Qy 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 799 AATGCCAATGATACGAATGCTCATGCTACGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 858
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 859 TCCACTAATAAGGAATGGGCTCAGGCGAATCTAGTCTTCCATCTATCATGATGATGAC 918
Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 919 GGTGGGGGAGCTTGGAGGACTTCCTCGAATCTCGAAGCTTATTCAGGCCAAGCATACAGT 978
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyValaTyrThr 140
Db 979 GCTGTGTCAGAAATTCATCAAACTCTGGGAGGACGAGTGAATGGGGCTTACCAACA 1038
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1039 GATTCAGAAATGAGTACATGATGTCGCAAAATGATATGACGATCTCTTTCGCTGCC 1098
Qy 161 GlyAsnGlyGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1099 GGAATGAAGGACCGGACCGGAGCCATCATGTCAGCCAGGACGACGCTAAATAATCAATA 1158
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1159 ACAGTCGGAGTACGAAACCTCCGCCCAAGCTTTGGTCTTATGCGGACATATCAAC 1218
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1219 CATGTGGCAGTCTCTTTCACGTGGACCGCAAAAGGATGGAGCGATCAAAACCGGATGTC 1278
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1279 ATGCCACCGGAAACGTTCTATCTATCAGCAAGATCTTCTTTCGACCGGATTCCTCTTC 1338
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
Db 1339 TGGCGGAACCATGACAGTAATATGACATCATGGTGGAAACGTCCTCATGCTACACCGATC 1398
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGlyHisPheValLysAsnArgGlyIleThrPro 280
Db 1399 GTTGCTGGAAACGTGCAAGCTTCGTGAGCATTTTGTGAAACACAGGAGCATCACACA 1458
Qy 281 LysProSerLeuLeuLysAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 1459 AAGCCTTCTCTATTAAGGCGGCACTGATTCGCGGTGACGTCGATCATGCGCTTGGCTAC 1518
Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValaTyr 320
Db 1519 CCGAAGCGTAACCAAGATGGGACGAGTGACATTTGGATAAATCCCTGAACGTTGCCCTAT 1578
Qy 321 ValAsnGlySerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 1579 GTGAACGAGTCCAGTCTCTATCCACACCAACCAAAAGGACGCTACTCGTTTACTGCTACT 1638

US-09-509-814A-7
; Sequence 7, Application US/09509814A
; Patent No. 6378227
; GENERAL INFORMATION:
; APPLICANT: TAKAIIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUMI
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; PRIORITY FILING DATE: 2000-04-06
; PRIORITY FILING DATE: 1998-10-07
; PRIORITY FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-7

Alignment Scores:
Pred. No.: 3.8e-224 Length: 1923
Score: 2242.00 Matches: 433
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 1
Query Match: 99.78% Indels: 0
DB: 4 Gaps: 0

US-09-985-689A-1 (1-434) x US-09-509-814A-7 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 619 AATGATGTTTGGCGTGGGATTTGTAAGCGGATGTTGGCTCAGACGAGCTACGGGTTGTAT 678
Qy 21 GlyGlnGlyGlnIleValAlaAspThrGlyLeuAspThrGlyLeuAspThrGlyLeuAspThr 40
Db 679 GGCAAGGACAGATCGTACGGTGGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 738
Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
```

739	TCGATCGCATGAAGCCCTTCGCGGGGAAAAATTACTGCGCAATTATATCGATTGGACGGACGAAT	798
61	AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly	80
799	AATGCCAATGATACGAATGCTCATGTGACGATGTGGCTGGCTCCGATTAGGAAACGGC	858
81	SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer	100
859	TCACATTAATAAAGGAATGGCGCTCAGGGGAAATCTTAGTCTTCCAAATCTATCATGGATAGC	918
101	GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer	120
919	GGTGGGGGACTTCGGAGGACTACCTCGAATCTCGAAACCTTATTCAGCCAGCATACAGT	978
121	AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr	140
979	GCTGGTGCCAGAATTTCATACAAATCTCGGGGAGCAGCATGTAATGGGGCTTACACAACA	1038
141	AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla	160
1039	GAITCCAGAAATGTGGATGACTATGTGGCCAAANAATGATATGACGATCTCTTTGGCTGCC	1098
161	GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle	180
1099	GGGAATGAAGGACCGGAACGGCGGAACCATCAGTGCACCGAGCAGACTAAAAATGCAATA	1158
181	ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn	200
1159	ACAGTCGGAGCTACGGAAAAACCTCCGCCCAAGCTTTGGGTCTTATCGGCAATATCAAC	1218
201	HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal	220
1219	CATGTGSCACAGTTCTCTTCAGCTGGACCGACAAAGGATGACGAGTCAAAACGGATGTC	1278
221	MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe	240
1279	ATGCAACCGGGAACGTTTCATACATCAGCAAGATCTCTCTTCGACCGGATTCCTCCTTC	1338
241	TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle	260
1339	TGGCGCAACCATGACAGTAANTATGCATACATGGGTGGAACGTCCTAGGCTACACCGATC	1398
261	ValAlaGlyAsnValAlaGlnLeuArgLysGluPheValLysAsnArgGlyIleThrPro	280
1399	GTTGCTGGAAACGTCGGACACAGCTCTGTGAGCATTTTGTGAAAAACAGAGCATCACACCA	1458
281	LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLysGlyTyr	300
1459	AAGCCTTCTTATTAAGAAGCGCACTGAATGCCGGTGCACTGACATCGGCCCTTGGCTAC	1518
301	ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr	320
1519	CCGAACGGTAACCAAGGATGGGACGAGTGACATTTGGATAAATCCCTTGAACGTTGCCTAT	1578
321	ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr	340
1579	GTGAACGAGTCCAGTTCTCTATCCACAGCAAAAGCAGCAGTACTCGTTTACTGTCTACT	1638
341	AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla	360
1639	GCGGGCAAGCCTTTGAAANAATCCCTCGTATGGTCTGATGCCCTCGGAGCAACAATGCT	1698
361	SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr	380
1699	TCCGTAAACGGTTGTCAATGATCTGGACCTTGTCAATTACCGCTCCAAATGGCACACAGTAT	1758
381	ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu	400
1759	GTAGGAATGACTTTACTTCCGCATACAATGATAACTGGGATGGCCCAATTAACGTAGAA	1818
401	AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn	420
1819	AATGTAATTAATGACCAAAAGCGGACGATATCAATTTGAATGATACAGGCTTAAC	1878

Qy 421 valProvalGlyProGlnThrPheSerLeuAlaIlevalAsn 434
|||
Db 1879 GTACCGGTTGGACCACAGAACTTCTCGTTGGCAATTCTTGAAT 1920

427

RESULT 3

```

US-09-509-814A-3
; Sequence 3, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUOASA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUMI
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A-3
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1920
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1920)
US-09-509-814A-3

```

Alignment Scores:			
Pred. No.:	5.47e-218	Length:	1920
Score:	2183.00	Matches:	418
Percent Similarity:	99.31%	Conservative:	13
Best Local Similarity:	96.31%	Mismatches:	3
Query Match:	97.15%	Indels:	0
DB:	4	Gaps:	0

US-09-985-689A-1 (1-434) x US-09-509-814A-3 (1-1920)

Qy	1	AsnAspValAlaArgGlyLeuValValAlaAspValAlaGlnSerSerTyrGlyLeuTyr	20
Db	616	AATGATGGCCAGAGGTATTGTCAAAGCGGATGTGGCAGAGCAGCTACAGTTTGAT	675
Qy	21	GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer	40
Db	676	GGCAAGGCCAGATTGTCCCGAGTTGCCGATCTGGATTGGATACAGAGAAACGACGT	735
Qy	41	SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn	60
Db	736	TCGATGCATGAAGCCTTCGCGGGTAAATAACAGCACTATATGCATCTGGTTCGACGAAT	795
Qy	61	AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly	80
Db	796	AATGGCAATGATACGACGGTCATGGTACCCATGTGCAGGTTCGGTATTAGGAATGCG	855
Qy	81	SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer	100
Db	856	GCAACGAATTAAGGAATGCCACCTCAAGCGCAATCTGGTTTTTCAAATCCATCATGGATAGC	915
Qy	101	GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer	120
Db	916	AGTGTGGGCTTGGAGGGTTGCCATCTCCAAATCTGCAAACCTTATTACGCCACGATTCAGT	975
Qy	121	AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr	140

Db 976 GCAGGTGCCAGAAATTCATACAACTCCTGGGGGGCAGCGGTGAATGGGGCTCACAGACA 1035
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1036 GATTCAGAAATGTGGTACATATGTAAGGAAAAATGATACAGATTCTTTTCGGCGCT 1095
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1096 GGGATGAAGCCCGACGCGGTACCATCAGTGCACTGGTACGGCTAAACCGCCATA 1155
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1156 ACAGTCGGCGCAACCGCAACCTGGCTGCAAGCTTCGGTTCTATGTCAGATAAATTAAC 1215
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyValGlyIleLysProAspVal 220
 Db 1216 CACGTTGCACAGTTCTTCCGTCGGCCGACAAAGATGGCGAATCAAGCCTGATGTC 1275
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 1276 ATGGCGCCAGGACATACATTTTATCAGCAAGATCTTCTTGCAACCGGATTCCTCTTC 1335
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 Db 1336 TGGCGGAATCATGACAGCAAAATATGCTATATGGTGGAGCTCATGCGCAACCGGAT 1395
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 1396 GTTGGGGGAATGTTGCACAGCTCGGTGAGCATTTTGTGAAATATAGAGGAATCACTCCT 1455
 Qy 281 LysProSerLeuLeuLysAlaIleAlaIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1456 AAGCCTTCCTTATGAACACAGCTTGTATGTCAGGTGCTGCTGATGTTGGATTTGGTTAT 1515
 Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1516 CCGAACGGAAACCAAGATGGGGCGGAGTGACCTGGATAAATCGTTGAACGTTGCCAT 1575
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 1576 GTGAACGAATCCAGTSCCTATCAATAGCCAAAAGCGACATATACCTTTACTGCAACG 1635
 Qy 341 AlaGlyProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 Db 1636 GCGGGCAACCAATCGAAATCTCCCTGGTATGGTGGATGCCCTGCGAAGCACTACTGCT 1695
 Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1696 TCTGTAACCTCGTCAATGATTGGATTTGGTCAITACAGCACCAACCGAACAAGATAT 1755
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyValArgAsnValGlu 400
 Db 1756 GTGGGAATGACTTTCACACCAATTTGCAATTAAGTGGATGGCGGCGCAATACGTAGAA 1815
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1816 AATGATTTATTAATTCGCCCCAAGTGAACATATACCATTTGAGGTGCAAGCATATAAT 1875
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1876 GTGCCGTTGGACCAAAACTTCTCGTTGGCAATTTGTAAC 1917

RESULT 4

US-08-873-479-41
 ; Sequence 41, Application US/08873479
 ; Patent No. 5891701
 ; GENERAL INFORMATION:
 ; APPLICANT: Slna, Alan
 ; APPLICANT: Lynne, Christanson
 ; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
 ; TITLE OF INVENTION: Having Protease Activity
 ; NUMBER OF SEQUENCES: 57
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5891701c No. 5891701disk of No. 5891701th America
 ; STREET: 405 Lexington Avenue
 ; CITY: New York
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10174
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/873,479
 ; FILING DATE: 12-JUN-1997
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Agtis, Cheryl H
 ; REGISTRATION NUMBER: 34,086
 ; REFERENCE/DOCKET NUMBER: 5251.000-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-867-0123
 ; TELEFAX: 212-878-9655
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 41:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3003 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; US-08-873-479-41

Alignment Scores:
 Pred. No.: 1,12e-211 Length: 3003
 Score: 2125.50 Matches: 406
 Percent Similarity: 97.93% Conservative: 19
 Best Local Similarity: 93.55% Mismatches: 8
 Query Match: 94.59% Indels: 1
 DB: 2 Gaps: 1

US-09-985-689A-1 (1-434) x US-08-873-479-41 (1-3003)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 1470 AATGACGTGCGCGTGGCATTTGAAAGCAGACGTCGCACAAATACCTTGGCTTATAT 1529
 Qy 21 GlyClnGlyClnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 1530 GGACAAAGGACAGATTGTAGCAGTTGCTGATCTGGGCTTGATACAGAAAGAAATGACAGT 1589
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 1590 TCGATGATGACAGCAATTCGCGGTAAAGATTACCGCACTATATGCACTGGGCAGACGAAT 1649
 Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 1650 AACGCCAATGATCAAAATGGACATGGAAACCCATGTTGCTGGATCTGTGTAGAAAT--- 1706
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 1707 GCTACAAATAAAGGATGGCAGCCGCAAGCAATCTAGTCTTTCAATCTATTATGATAGT 1766
 Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 1767 GGTGGAGGCTGGGAGGACTACCTGCTAATCTCAAAACATTTATTCAGTCAAGCATATAGT 1826
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValAlaValAsnGlyAlaTyrThrThr 140
 Db 1827 GCTGGAGCGAAGATTCATCAAGATTCATGGGGGCTCCAGTAAACGGTGCCTATACGACA 1886
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1887 GACTCTCGAAATGTTGATGATTATGTGAGAAAAAATGATATGACGATCTTTTGGGGCC 1946

161 GlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1947 GGAATGAGGAGCAGATGAGTACATCATCGTCCAGGACGACGACGACGACGAT 2006
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 2007 ACAGTTGGGGCAACCGGAAACCTACGTCGAAGCTTCGATCTTATCGGATATATTAAC 2066
Qy 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyValArgIleLysProAspVal 220
Db 2067 CATGTTGCTCAATTTCTTCCAGAGGTCCTACAGAGTGGAGTATTAAGCCGACGTC 2126
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 2127 ATGGCACCAGTACGATATTTCTCTGCTAGATCATCATTAGCTCCAGATTCCTCATTC 2186
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
Db 2187 TGGCAAAACCATGATGATTAATATGCTACATGGGTGTTACTTCTATGGCTACTCCAAT 2246
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 2247 GTACAGGTAATGTGCACATTAAGGAGCATTTTGTGAAATATAGAGGGTAACTCCT 2306
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 2307 AAGCCTTCCCTTTTAAAGCTGCTTTAATTTGAGGTGCTCGGATGTGGACTTGGCTTT 2366
Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 2367 CCATGTTACCAAGAGTGGGAGAGTAACTGTTAGTAAATCCCTTAATGTCCGATTT 2426
Qy 321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 2427 GTGAATGAACAGAGCCCTTTTCAACAAAGTCAAAAGACACATATTCGTTACGGCTCAA 2486
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrAla 360
Db 2487 GCTGGTAACCTTTAAATAATACATTTGTTGTCAGATGCACAGTAGCAGCGGCA 2546
Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 2547 TCATAACTTTAGTGAATGATTAGACTTAGTAACTCACTGACCAACAAATGAATAATAC 2606
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 2607 GTCCGAAATGACTTTACAGCACCCTGATGATAACAATTTGGGATGGCAGAAACACGTTGAA 2666
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 2667 AATGTGTTTATCAATGCTCTCTCAAGCGGAACGATATACATCGAAGTGCAGGCTTACAT 2726
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 2727 GTACCAATATTCGCAACCTTTCTTTTACGATTTGATAT 2768

RESULT 5
US-08-894-818B-2
; Sequence 2, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masenori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Kunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/894,818B
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1977 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-2

Alignment Scores:
Pred. No.: 2,57e-37 Length: 1977
Score: 452.50 Matches: 138
Percent Similarity: 44.66% Conservative: 67
Best Local Similarity: 30.07% Mismatches: 153
Query Match: 20.14% Indels: 101
DB: 3 Gaps: 18

US-09-985-689A-1 (1-434) x US-08-894-818B-2 (1-1977)

Qy 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnIleValAla 27
Db 433 ATAGGGGCGGATACCGTCTGGAACTCCTCGGCTACGACGAGCGGTGTGGTGGTCC 492
Qy 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATCTGCTGATACGGGTATAGACGCGAAC-----CACCCCGATCTGAAG 534
Qy 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
Db 535 GCGAAGGTATAGGCTGTGTACGCGCGCTCAACGCGAGTCCGCCCTCTACGTACGACAG 594
Qy 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
Db 595 GGACACGGAAACCCAGTTCGCGGTATCGTTGCCGGAACCGGCGCGGTAACTCCAGTAC 654
Qy 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
Db 655 ATAGCGTCCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 714
Qy 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
Db 715 GGAACGCTTCACCATCATCTCGCGGTGTGACTGGGTCTGTCAGAACACGACGACGATAC 774
Qy 122 GlyAlaArgIle-----HisThrAsnSer 129
Db 775 GGGATAAGGGTCATCAACCTCTCCTCGGCTCCTCCAGAGCTCCGACGAGAACCGACTCC 834

```

QY 130 TrpGlyAlaValAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyrVal 149
DB 835 CTCAGTCAGCGCGTCAACAGCGCTGGGACGC----- 867
QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyThr 169
DB 868 -----GGTATAGTAGTCTGCGTCCGCGCGCAACAGCGCGCGCAACACCTACAC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
DB 919 GTGGCTCACCCGCCCGCGAGCAAGTCAACCGTCGGTGCA----- 963
QY 190 ProSerPheGlySerTyrAlaAsnAsnIleAsnHisValAlaGlnPheSerArgGly 209
DB 964 -----GTTGACAGCAACAGCAACATCGCAGCTTCTCCAGCAGGGGA 1005
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
DB 1006 CCAGCCGCGAGGAAGGCTCAAGCGGAAGTGTGCGCCCGCGGTGACATCATAGCC 1065
QY 230 AlaArgSerSerLeuAlaProAspSerPheTyrAlaAsnHisAspSerLysTyrAla 249
DB 1066 CCGCGCGCCAGC-----GGAACAGCATGGGCACCCCGATAAACGACTACTACAC 1116
QY 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu 268
DB 1117 AAGCGCTCTGAACCAAGCATGGCCCGCGACGTTTCGGCGGTGGCGCGCTCATCTC 1176
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
DB 1177 CAGGCCAC-----CCGAGTGGACCCCGGACAAAGTGAAG 1212
QY 287 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
DB 1213 ACCGCCCTCATCAGACCCCGCATAGTCGCCCCCAAGAGATAGCGGACATCGCTAC 1272
QY 299 GlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeu----- 316
DB 1273 GGTGCG-----GGTAGGTGAACGTCTACAAGGCCATCAAGTAC 1311
QY 317 ---AsnValAlaTyrValAsnGlnLysSerSerLeuSerThrSerGlnLysAlaThrTyr 335
DB 1312 GAGGACTACCCCAAGCTCACCTTCACCGGCTCGTCGCGCAACAGGGGAAGCGCCAC 1371
QY 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaPro 355
DB 1372 ACCTTCAGCTAGCGCGCCACCTTCGTCGACCGCCACCTCTACTGGGAC----- 1422
QY 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAspValIleThrAlaPro 375
DB 1423 -----ACGGGCTCGAGGACATCGACCTCTACCTCTACGACCC 1461
QY 376 AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGly 395
DB 1462 AACGGGAACGAG---GTGACTACTCTACCGCCCTACTAC----- 1500
QY 396 ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrThrIleGlu 415
DB 1501 -----GGCTTCGAGAAGGTGCGCTACTACAACCCGACCGCGGAACCTGCGGTC 1554
QY 416 ValGlnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
DB 1555 GTCGTCAGCTACAAG-----GGCGGCGGAACCTACACGAGTCGACGTCGTCAGC 1602

```

RESULT 6

```

US-09-445-472-11
; Sequence 11, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko

```

```

; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1377

```

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-11

```

```

Alignment Scores:
Pred. No.: 2,57e-37 Length: 1977
Score: 452.50 Matches: 138
Percent Similarity: 44.66% Conservative: 67
Best Local Similarity: 30.07% Mismatches: 153
Query Match: 20.14% Indels: 101
DB: 4 Gaps: 18

```

US-09-985-689A-1 (1-434) x US-09-445-472-11 (1-1977)

```

QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAla 27
DB 433 ATAGGGCCGATACCGTCTGGAACCTCCCTCGGTACGACGGAAGCGGTGTGGTGGTGC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
DB 493 ATCTGCTAGCGGTATAGACCGGAAC-----CACCCGATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
DB 535 GGCAAGGTCATAGCTGGTACGACGCGCGTCAACGGCAGGTGCGACCCCTACGATGACCAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
DB 595 GGACACGAAACCCACGTTGCGGTATCGTTCGCGGAACCGGACGCTTAATCTCCAGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
DB 655 ATAGCGTCGCGCCCGCGCGAAGCTCGTCGGGTCAAGGTTCTCGGTGCGGACGCTTCG 714
QY 102 GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
DB 715 GGAAGCGTCTCCACCATCATCGCGGGTGTGACTGGGTCTCCAGAACAAAGGCAAGTAC 774
QY 122 GlyAlaArgIle-----HisThrAsnSer 129
DB 775 GGGATAGGTCATCACTCTCCCTCGGCTCTCCAGAGCTCCGACGGAACCGACTCC 834
QY 130 TrpGlyAlaAlaValAsnGlyAlaTyrThrAspSerArgAsnValAspTyrVal 149
DB 835 CTCAGTCAGCGCGTCAACAACGCTGGGACGCC----- 867
QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr 169
DB 868 -----GGTATAGTAGTCTGCGTCCGCGCGCAACAGCGCGCGCAACCTACAC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
DB 919 GTCGCTCACCCCGCGCGGAGCAAGTCAATACCGTCGCTGCA----- 963
QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
DB 964 -----GTTGACAGCAACAGCAACATCGCAGCTTCTCCAGCAGGGGA 1005
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229

```


QY 294 -----AlaAspIleGlyLeuGlyTyProAsnGlyAsnGlnGlyTrpGlyArgVa 310
 Db 852 AGATGAATAGCGGATATAGCTACGGTGCA-----GGTAGGT 890
 QY 310 lThrLeuAspIleGlyLeuValAlaTyValAsnGlnGlySerSerLeuSerThrSe 330
 Db 891 TAATGCATCAAGGCTATAAAC-----TACGATAACTATGCAAGCTAGTGTCACTGG 944
 QY 330 rClnIlySala-----ThrTySerPheThrAlaThrAlaGlyLysProLe 345
 Db 945 ATATGTTGCCAACAAAGGCGCAACTCACCAGTTCGTTATTAGCGAGGTTCTGCTGCT 1004
 QY 345 uIlySileSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 Db 1005 AACTGCCACATATATCGGCAATGCCAAT----- 1035
 QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyValGlyAsnAspPh 385
 Db 1036 AGCGACCTTGATCTTACCTCTACGATCCCAAGCAACAGC---GTGACTACTCTTA 1091
 QY 385 eThrSerProTyAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
 Db 1092 CACCGCTACTAT-----GGATTGGAAGGTTGGTTATTA 1127
 QY 405 nAlaProGlnSerGlyThyThyThrIleGluValGlnAlaTyAsnValProValGlyPr 425
 Db 1128 CAACCCAACTGATGGAACATGGAATTAAGGTTGTAAGCTACAGC-----GGAAG 1178
 QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
 Db 1179 TGCMAACTATCAAGTAGATGTGTTAAGT 1206

RESULT 8

US-08-894-818B-4
 ; Sequence 4, Application US/088948:8B
 ; Patent No. 6261822
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: YAMAMOTO, Katsuhiko
 ; APPLICANT: MITTA, Masanori
 ; APPLICANT: ASADA, Kiyozo
 ; APPLICANT: TSUNASAWA, Susumu
 ; APPLICANT: KATO, Ikunoshin
 ; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Browdy and Neimark
 ; STREET: 419 Seventh Street N.W., Ste. 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: United States of America
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/894,818B
 ; FILING DATE: 20-MAY-1998
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/JP96/03253
 ; FILING DATE: 07-NOV-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 323285/1995
 ; FILING DATE: 12-DEC-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Browdy, Roger L.
 ; REGISTRATION NUMBER: 25,618
 ; REFERENCE/DOCKET NUMBER: TAKAKURA=1

TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 628-5197
 ; TELEFAX: (202) 737-3528
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1566 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: genomic DNA
 ; FEATURE:
 ; OTHER INFORMATION: /note= N at position 1283 is G or T.
 ; US-08-894-818B-4

Alignment Scores:

Pred. No.: 1,02e-33 Length: 1566
 Score: 416.50 Matches: 141
 Percent Similarity: 42.55% Conservative: 59
 Best Local Similarity: 30.00% Mismatches: 148
 Query Match: 18.54% Indels: 122
 DB: 3 Gaps: 20

US-09-985-689A-1 (1-434) x US-08-894-818B-4 (1-1566)

QY 12 ValAlaGlnSerSerTyGly-LeuTy-----GlyGlnGlyG1 24
 Db 24 GTCTGCAGCTCAAGTATATGCAACTTACGTTTGGAACTTGGGATATGATGGTCTGGAAT 83
 QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMethHisG1 44
 Db 84 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 125
 QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 Db 126 AGATCTCCAAGGAAGAAGTA-----ATTGGTGGGTAGATTTTGTCAATGG 170
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
 Db 171 TAGGAGTTATCCATACGATGACCATGACATGGAACATCATGTAGCTTCAATAGCAGCTGG 230
 QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 Db 231 TACTGGAGCAGCAAGTAATGGCAAGTACAGGAATGGCTCCAGGAGTAAGTGGCGGG 290
 QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
 Db 291 AATTAAGGTTCTAGGTCCCGATGCTTCTGGAACATATCTACTATAATTAAGGAGCTTGA 350
 QY 112 nThrLeuPheSerGlnAlaTySerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
 Db 351 GTGGCGGTTGTATAACAAGATAAGTACGGAATTAAGTCAATTAATCTTCTTCTGTTTC 410
 QY 132 a-----AlaValAsnGlyValaTyThrTh 140
 Db 411 AAGCCAGAGCTCAGATGGTACTACGCTCTAAGTCAGGCTGTTAATGCGGCGGATGC 470
 QY 140 rAspSerArgAsnValAspAspTyValArgLysAsnAspMethTrileuPheAlaAl 160
 Db 471 T-----GGATTAGTTGTTGTGTTCCCGC 494
 QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla11 180
 Db 495 TGGAAACAGTGGACCTAACAGTATACATCGGTTCTCCAGCAGCTGCAAGCAAGTTAT 554
 QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAs 200
 Db 555 TACAGTTGGAGCC-----GTTGACAAAGTATGA 591
 QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 Db 582 TGTATTAAACAAGCTTCTCAAGCAGAGAGGCGCAATGACAGCGGAGGCTTAAGCTGAGGT 641
 QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240

```
Db 642 TGTGCTCCAGAACTGATAATTCTCCAGCAAGT-...-GGAACTAGCAT 692
QY 240 eTrpAlaAsnHisAspSerLysTyAlaTyMetGlyGlyThrSerMetAlaThrProIl 260
Db 693 GGGTCAACCAATTAAGTACTATTACACAGCAGCTCTGGACATCAATGGCACTCTCA 752
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeuThrPr 280
Db 753 CGTAGCTGTATTGCAAGCCCTCTGTCCAA-...-GCACACCC 791
QY 280 olys-...-ProSerLeuLeuLysAlaAlaLeuLeuAlaGlyAla-...- 293
Db 792 GAGCTGGATCCAGCAAGAAAGAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 851
QY 294 -...-AlaAspIleGlyLeuGlyTyProAsnGlyAsnGlnGlyTyTrpGlyArgVa 310
Db 852 AGATGAATAGCGCATATAGCTTACCGTGCA-...-GGTAGGT 890
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyValAsnGluSerSerSerLeuSerThrSe 330
Db 891 TAATGATACACAGCTATAAAC-...-TACGATAACTATGCAAGCTAGTCTTCACTGG 944
QY 330 rGlnLysAla-...-ThrTySerPheThrAlaThrAlaGlyLysProLe 345
Db 945 ATATGTTGCCAACAAAGGAGCCAAACTCACAGTTCGTTATTAGCGAGCTTCGTTCT 1004
QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1005 AACTGCCACATATATCTGGACAACTCCAA-...- 1035
QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyValGlyAsnAspPh 385
Db 1036 -AGCGACCTTGATCTTACCTCAGTCCCAATCCATGGAACCCAG-...-GTTGACTACTCTTA 1091
QY 385 eThrSerProTyAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1092 CACCGCTACTAT-...-GGATTCGAAAGGTTGGTTATTA 1127
QY 405 nAlaProGlnSerGlyThrTyThrIleGluValGlnAlaTyAsnValProValGlyPr 425
Db 1128 CAACCCCACTGATGGACATGGAATTAAGTTGTAAGCTACAGC-...-GGAG 1178
QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTATCAAGTAGATGTGTAAGT 1206
```

RESULT 9

```
US-08-894-818B-34
; Sequence 34, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA: JP 323285/1995
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1962 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-894-818B-34
; Alignment Scores:
; Pred. No.: 1,46e-33 Length: 1962
; Score: 416.50 Matches: 141
; Percent Similarity: 42.55% Conservative: 59
; Best Local Similarity: 30.00% Mismatches: 148
; Query Match: 18.54% Indels: 122
; DB: 3 Gaps: 20
; US-09-985-689A-1 (1-434) x US-08-894-818B-34 (1-1962)
QY 12 ValAlaGlnSerSerTyGly-LeuTy-...-GlyGlnGlyG1 24
Db 420 GTCTGCAGCTCAAGTTATGCGAATTCAGTTTGGAACTGGGATATGATGGTTCTGGAAT 479
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 480 CACAATAGGAATAATTGCACACTGGAATTGAC-...-GCTTCTCATCC 521
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsn- 63
Db 522 AGATCTCCAGGAAAGTA-...-ATTGGGTGGGTAGATTGTTCATGG 566
QY 64 -...-AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 567 TAGAGTTATCCATGACATGACATGACATGACATGACATGACATGACATGACATGACATG 626
QY 78 yAsnGlySerThrAsn-...-LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGCAAGTACAAGGAATGCTCCAGGAGCTAAGCTGGCGGG 686
QY 94 eGlnSerIleMet-...-AspSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
Db 687 AATTAAGGTTCTAGTCCCGGATGTTCTGGAGCATATCTACTATATAATTAAGGAGTTGA 746
QY 112 nThrLeuPheSerGlnAlaTySerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGCCCGTTGATAACAAAGATAAGTACGGAATTAAGTCATTAATCTTCTCTGTGCTC 806
QY 132 a-...-AlaValAsnGlyAlaTyThrTh 140
Db 807 AAGCCAGAGCTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 866
QY 140 rAspSerArgAsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 T-...-GGATTAGTCTGTGTGGTGGTGGC 890
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
```

```
Db 891 TGGAAACAGTGGACCTAACAGATATACAAATCGTTCTCCAGCAGCTGCAACCAAGTTAT 950
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerThrAlaAspAsnIleAs 200
Db 951 TACAGTTGGAGCC-----GTGCAACAGTATGA 977
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 978 TGTGTATACAAAGCTTCTCAAGCAGAGGCGCACTGACAGCGCAGGCTTAAGCTGAGGT 1037
Qy 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTGTCTCCAGAACTGGATAATGCTGCCAGACAAAGT-----GGAACCTAGCAT 1088
Qy 240 eThrAlaAsnHisAspSerLysThrAlaThrMetGlyGlyThrSerMetAlaThrProIl 260
Db 1089 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTGGACATCAATGGCACTCTCA 1148
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 1149 CGTAGCTGTATTGCGAGCCCTCTGCTCCAA-----GCACACCC 1187
Qy 280 olys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 1188 GAGCTGGACTCCAGCAAAAGTAAACAGCCCTCTATAGAACTGCTGATATCGTAAAGCC 1247
Qy 294 -----AlaAspIleGlyLeuGlyTyPrProAsnGlyAsnGlnGlyTyPrGlyArgVa 310
Db 1248 AGATGAATAGCGCATATAGCTTACCGTGCA-----GGTAGGGT 1286
Qy 310 lThrLeuAspLysSerLeuAsnValAlaThrValAsnGlnSerSerSerLeuSerThrSe 330
Db 1287 TAATGCATACAGGCTATAAAC-----TACGATTAACATATGCAAGCTAGTGTTCACCTGG 1340
Qy 330 rGlnLysAla-----ThrTySerPheThrAlaThrAlaGlyLysProLe 345
Db 1341 ATATGTTGCCAAACAAAGGCGAGCAACCTCACCAGTTGTTATTAGCGGAGCTTCGTTCT 1400
Qy 345 uLysIleSerLeuValThrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATTATATCGGCAATGCCAAT----- 1431
Qy 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyValGlyAsnAspPh 385
Db 1432 -AGCGACCTTGATCTTACCTCTACGATCCCAATGGAAACCAAG--GTTGACTACTCTTA 1487
Qy 385 eThrSerProTyAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1488 CACCGCTACTAT-----GGATTGCAAAAGGTTGGTTATTA 1523
Qy 405 nAlaProGlnSerGlyThrTyThrIleGluValGlnAlaThrAsnValProValGlyPr 425
Db 1524 CAACCCACATGTGGAAACATGACATTAAGTTGTAGCTACAGC-----GGAAG 1574
Qy 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1575 TGCAAACTATCAAGTAGATGTGGTAGT 1602
```

RESULT 10

```
US-09-445-472-15
; Sequence 15, Application US/09445472
; Patent No. 6356726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikumoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
```

```
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-15

Alignment Scores:
Pred. No.: 1,46e-33 Length: 1962
Score: 416.50 Matches: 141
Percent Similarity: 42.55% Conservative: 59
Best Local Similarity: 30.00% Mismatches: 148
Query Match: 18.54% Indels: 122
DB: 4 Gaps: 20
```

US-09-985-689A-1 (1-434) x US-09-445-472-15 (1-1962)

```
Qy 12 ValAlaGlnSerSerTyGly-LeuTy-----GlyGlnGlyG 24
Db 420 GTCTGCAGCTCAAGTTATGCAACTTACGTTTGGAACTTGGGATATGATGTTCTGGAAT 479
Qy 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMethIsl 44
Db 480 CACAATAGGAATAATTGACACTGGAATGAC-----GCITCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAAGGAAAGTA-----ATTGGTGGGTAGATTGTTCAATGG 566
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db 567 TAGGAGTTATCCATACGATGACCATGACATGCACTGCACTCATGTAGCTCAATAGCAGCTG 626
Qy 78 YAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGACGACGAGTAATGGCAAGTACAAAGGATGGTCCAGGAGCTTAAGCTGCGGG 686
Qy 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGl 112
Db 687 AATTAAGGTTCTAGTGCCGATGGTCTTGGAAACATATCTACTATAATTAAAGGAGTGA 746
Qy 112 nThrLeuPheSerGlnAlaThrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGGCGGTTGATAACAAGATAAGTACGGAATTAAGTCATTAACTTCTTCTTCTGTTTC 806
Qy 132 a-----AlaValAsnGlyAlaTyThrTh 140
Db 807 AAGCCAGAGCTCAGATGGTACTGACGCTCTTAAGTCAGGCTGTTTAATGACGCGTGGATGC 866
Qy 140 rAspSerArgAsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 T-----GGATTAGTTGTTGTGTTGTTCCCGC 890
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
Db 891 TGGAAACAGTGGACCTTAACAAGTATACAATCGGTTCTCCAGCAGCTGCAAGCAAGTTAT 950
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyThrAlaAspAsnIleAs 200
Db 951 TACAGTTGGAGCC-----GTTGACAAAGTATGA 977
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 978 TGTGTATACAAAGCTTCTCAGCAGAGAGGCGCACTGACAGCGCAGGCTTAAGCTGAGGT 1037
Qy 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTGTCTCCAGAACTGGATAATGCTGCCAGACAAAGT-----GGAACCTAGCAT 1088
```

```
QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrPro11 260
Db 1089 GGTACCAACAAATTAAGTACATACACAGCAGCTCTGGACATCAATGGCACTCTCTCA 1148
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 1149 CGTAGCTGTTAGTTCAGCCCTCTTGCTCAA-----GCACACCC 1187
QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuLeuAlaGlyAla----- 293
Db 1188 GAGCTGGAGCTCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247
QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
Db 1248 AGATGAATAGCGGATATAGCTACGGTCA-----GGTAGGGT 1286
QY 310 ThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
Db 1287 TAATGCATCAAGGCTATAAAC-----TACGATTAATGCAAGAGCTAGTGTTCACCTGG 1340
QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
Db 1341 ATATGTTGCCAACAAAGGCGCAAACTCACCAGTTTCGTTATTAGCGGAGCTTCGTCGT 1400
QY 345 uLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATTAATCTGGGCAATGCCAAT----- 1431
QY 365 LAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
Db 1432 -ACGGACCTTGATCTTACCTACATGCCAATGGAACAG--GTTGACTACTCTTA 1487
QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1488 CACGGCTACTAT-----GGATTGAAAGGTTGTTATTA 1523
QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1524 CACCCCACTGATGGACATGGACATTAAGTTTGAAGCTACAGC-----GGAAG 1574
QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1575 TGCAAACTATCAAGTAGATGTGGTAAGT 1602

RESULT 11
US-08-894-818B-6
; Sequence 6, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikumoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
```

```
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: PCT/JP96/03253
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA: JP 323285/1995
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-894-818B-6

Alignment Scores:
Pred. No.: 3,36e-32 Length: 1977
Score: 403.50 Matches: 134
Percent Similarity: 43.07% Conservative: 68
Best Local Similarity: 28.57% Mismatches: 147
Query Match: 17.96% Indels: 120
DB: 3 Gaps: 20

US-09-985-689A-1 (1-434) x US-08-894-818B-6 (1-1977)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 420 GTCTGCAGCTCAAGTTATGCACTTACCTTGGAACTGGATATGATGTTCTGGAAT 479
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 480 CACATAGGAATATTGACACTGGAATTGAC-----GCTTCTCATCC 521
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAGGAAAGTA-----ATTGGGTGGGTAGATTTTGTCAATGG 566
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 567 TAGGAGTTATCCATACGATGACATGACATGGAACTCATGTAGCTTCAATAGCAGCTGG 626
QY 78 YAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGCAAGTACAAAGGAATGGCTCCAGGAGCTAAGCTGGCGGG 686
QY 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
Db 687 AATTAAGGTTCTTAGTGGCCGATGGTCTGGAAGCATCTACTACTAATTAATTAAGGAGTGA 746
QY 112 mThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle----- 125
Db 747 GTGGGCCGCTTGATAACAAAGATAAGTAGCGGAATTAAGGTCAATCTTCTCTGGTTC 806
QY 126 -----HisThrAsnSerTrpGlyValAlaValAsnGlyAlaTyrThrTh 140
Db 807 AAGCCAGAGCTCCGACGGAACCGACTCCCTCAGTCAGCCGCTACACACGCTCGGACGC 866
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 C-----GGTATAGTAGTCTCGCTCGCGCGC 890
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 180
Db 891 CGGCAACACGCGCGCGCAACACCTACACCGTCGGCTCACCCCGCCCGCGAGCAAGGTCTAT 950
```

```
QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 951 AACCGTCGGTGA-----GTTGACGACGACGA 977
QY 200 nHisValAlaGlnPheSerArgGlyProThrIleAspGlyArgIleIysProAspVa 220
Db 978 CAACATCCCGAGCTTCTCCAGCAGGGGACCGCGGACCGGAGGCTCAAGCGGAGT 1037
QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPh 240
Db 1038 GTGCGCCCGCGGTGATCATAGCCCGCGGCCAGC-----GGAACCATCAT 1088
QY 240 eTrpAlaAsnHisAspSerIysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIl 260
Db 1089 GGGCACCCCGGATAAACGACTACTACCAAGAGCCCTCTGAACACGACATGGCCACCCGCA 1148
QY 260 eValAlaGly---AsnValAlaGlnLeuArgGluHiePheValIysAsnArgIleTh 279
Db 1149 CGTTTCGGGGGTGGCGGCTCATCTCCAGGCCAC-----1185
QY 279 rProLys-----ProSerLeuLeuIysAlaAlaLeuIleAlaGlyAla----- 293
Db 1186 -CCGAGCTGGACCCCGGACCAAGGTGAAGACCGCCCTCATCGAGACCGCGACATAGTCG 1244
QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyAr 309
Db 1245 CCCCAGGAGATAGCGGACATCGCTACGGTGGC-----GGTAG 1283
QY 309 gValThrLeuAspIysSerLeu-----AsnValAlaTyrValAsnGluSerSe 326
Db 1284 GTGAACTGTACAGGACCATCAAGTACGACGACTACGCCAAGCTCACCTTCACCGGCTC 1343
QY 326 rLeuSerThrSerGlnIysAlaThrTyrSerPheThrAlaThrAlaGlyIysProLeuLy 346
Db 1344 CCGCCGACCAAGGAAGCGCCACACCTTCAGCTCAGCGCGGCCACCTTCGTGAC 1403
QY 346 sIleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuValAs 366
Db 1404 CGCCACCTCTACTGGAC-----ACGGCTCGAG 1433
QY 366 nAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPheTh 386
Db 1434 CGACATCGACCTCTACCTCTACGACCCCAACGGGAACGAG---GTTGACTACTCTCTAC 1490
QY 386 rSerProTyrAsnAspAsnTyrAspGlyArgAsnAsnValGluAsnValPheIleAsnAl 406
Db 1491 CCGCTACTAC-----GGCTTCGAGAGGTCGCTACTACAA 1526
QY 406 aProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyProGl 426
Db 1527 CCGGACCGCGGAACTGTGACGCGTCAAGTGTCTAGCTACAAG-----GGCGGGC 1577
QY 426 nThrPheSerLeuAlaIleValAsn 434
Db 1578 GAACACTACAGGTCGACGTCGTGACG 1602

RESULT 12
US-09-000-016-3
; Sequence 3, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
```

```
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
```

```
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
US-09-000-016-3
```

```
Alignment Scores:
Pred. No.: 1,51e-26 Length: 2539
Score: 351.00 Matches: 131
Percent Similarity: 42.63% Conservative: 54
Best Local Similarity: 30.18% Mismatches: 159
Query Match: 15.62% Indels: 90
DB: 3 Gaps: 16
```

US-09-985-689a-1 (1-434) x US-09-000-016-3 (1-2539)

```
QY 2 AspValAlaArgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACACGTCTGTCGGGAGATCGCGCCCCCAAGGCGTGTCTCGCC---GGCTACGACGCG 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
Db 1010 AAGGCGTGAAGATCGCGTCTCTGGACACCGGTGTCGACACGAGC-----1054
QY 42 MetHisGluAlaPheArgGlyIysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGGACCTGAAGGCGCGGGTGACCGGTCCAGAACTTACCCTCCGCGCGCGCG 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCCGGGACAGAGGTGGGCGCACCCACGTCGCTCGATCGCGCGCGGCGCGCGCGCG 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAGGCGAAGTACAGGCGCTCGACCCCGCGCGCGCGATCTCAACGCGAAGGTC 1231
QY 98 MetAspSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTGACGACTCCGCT-----TTGCGCGGACGACTCCGGCATCTCCGCGCGCATGGAGT 1285
QY 118 AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaValAlaValAsnGlyAla 137
```

Db 1286 GCGCCCGCGAGGCGCGCGCTGCTCACCATGAGCTGGC-----GGCATG 1333
Qy 138 TyrThrAspSerArgAsnValAspTyrValArgLys-----AsnAspMetThr 155
Db 1334 GACACCGGAGACCGCGCTGGAGCGCGCTGACAAAGCTGCGCGCGAAGAGGC 1393
Qy 156 IleLeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyThrIleSerAlaPro 173
Db 1394 GTCCTGTTCCTCCATCGCGCGCGCGCGCGCGCGCGAG-----TCGATCGGTTCGCC 1447
Qy 174 GlyThrAlaLysAsnAlaIleThrValGlyValAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GGCAGCGCGGAGCGCGCGCTCACCGTGGCGCC----- 1480
Qy 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro-----ThrLys 212
Db 1481 -----GTCGACGACAAAGCAAGCTCGCGACTTCCTCCACCGCGCGCGCTCGGC 1534
Qy 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
Db 1535 GACGCGCCATCAAGCCGCGAGCTCACCGCTCCCGCGGTGACATCAGCGCGCTCGGC 1594
Qy 233 SerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGly 252
Db 1595 GAGGCGCAACGACATCGCGCGAGGCTCGTGAGGCGCGCGCTACATGACCATCTCC 1654
Qy 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
Db 1655 GGCAGCTGATGGCGAGCCCGCAGCTCGCGCGCGCGCGCGCTCCGTAAGCAGCAG--- 1711
Qy 273 ValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
Db 1712 -----CACCCGCGACTGGACCTCCCGCGAAGTGAAGGCGCG 1747
Qy 290 IleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArg 309
Db 1748 CTCACCGGCTCCACCAAG---GGCGCAAGTACACCCCGTTCGAGCAGAGGTTCGGCGCG 1804
Qy 310 ValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerLeuSer--- 328
Db 1805 ATCCAGCGCGACAGCGCTCCAGCAGACCGGTATCCCGCGCGCGCTCGGTGAGCTTC 1864
Qy 329 -----ThrSerGlnLysAlaThrTyrSer 336
Db 1865 GCGCTCCAGCAGTGGCGCGCACACGACGACGCGCGGTCCACCAAGCAGCTGACCTACCGC 1924
Qy 337 PheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1925 AACCTCGGCACCCAGGAGCGTACGCTGAGCTGAGCTGACCGCGCGCGCGCGCGCGCG 1984
Qy 354 -----AlaProAlaSer 357
Db 1985 AAGCGCGCGCGCGCGGCTTCTTACGCTGGCGCGCCACCGGTGACCGTCCCGCGCGCG 2044
Qy 358 ThrThrAlaSerValThrLeuValAsnAspLeuAspLeu----- 370
Db 2045 GGCAGCGCGCTCCGTCGACATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2104
Qy 371 -----ValIleThrAlaProAsnGlyThrGlnTyrVal 381
Db 2105 TACTCGCGGTACGTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTC 2146

RESULT 13

US-09-514-340-3
; Sequence 3, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: AKIRA ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPIRIDINE DERIVATIVES
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-514-340-3

Alignment Scores:
Pred. No.: 1,51e-26 Length: 2539
Score: 351.00 Matches: 131
Percent Similarity: 42.63% Conservative: 54
Best Local Similarity: 30.18% Mismatches: 159
Query Match: 15.62% Indels: 90
DB: 4 Gaps: 16

US-09-985-689A-1 (1-434) x US-09-514-340-3 (1-2539)

Qy 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACACGTCGTCGCGCAGATCGCGCCCAAGGCGTGTCCGCC---GGCTACGACGCG 1009
Qy 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
Db 1010 AAGGCGGTGAAGATCGCGCTCTGGACACCGGTGTGACACGAGC----- 1054
Qy 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGGACCTGAAGGCGCGGTGACCGGTCCAGAACTTCACCGCGCGCGCGC 1111
Qy 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCGCGCGACAGGTGGCGCGCGCGCACCCACGCTCGCTCGATCGCGCGCGCGCGCGCC 1171
Qy 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAAGGCAAGTACAGGCGGTGCGACCCCGCGCGCGGTCTCTCAACGGCAAGGTC 1231

Qy	98	MetAspSerGlyGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln	117
Db	1232	CTCGAGCACTCCGGT-----TTCCGGCGACGACTCCGGCATCTCCCGCGCATGGAGTGG	1285
Qy	118	AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValaAsnGlyAla	137
Db	1286	CGCGCCGGCAGCGCGCGACGTCGTCCACCATGAGCTGGGC-----GGCATTG	1333
Qy	138	TyrThrThrAspSerArgAsnValAspAspTyrValArgLys-----AsnAspMetThr	155
Db	1334	GACACACCGGAGACCGACCCGCTGGAGCGCGCGTGCACAAAGTGTCGCCCGAGAGGCG	1393
Qy	156	IleLeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaPro	173
Db	1394	GTCCGTGTTCCATCGCGCGCGCAACGAGGGCCCGAG-----TCGATCGGTTCGCC	1447
Qy	174	GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly	193
Db	1448	GGCAGCGGACGCCCGCCCTCACCGTCGGCGCC-----	1480
Qy	194	SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro-----ThrLys	212
Db	1481	-----CTCAGCAGACGAGCAAGCTCGCGCACTTCTCTCCACCGGCCCCCGCTCGGC	1534
Qy	213	AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer	232
Db	1535	GACGGCCCATCAAGCCGGACGTCACCGCTCCCGCGCTGACATCAACGGCGCTCGGG	1594
Qy	233	SerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGly	252
Db	1595	GAGGGCAACGACATCGCCAGGAGTCCGTGAGGACCGCGCGCTACATGACCACTCC	1654
Qy	253	GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe	272
Db	1655	GGCACGTGATGGGACCCCGACGTCGCGGGCGCGCCCTCTCTGAAGCAGCAG--	1711
Qy	273	ValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----	289
Db	1712	-----CACCCGCACTGGACCTCCGCGCACTGAAGGCGCG	1747
Qy	290	IleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArg	309
Db	1748	CTCACCGGCTCCACCAAG--GGCGGAAGTACACCCCGTTCGAGCAGGCTTCGGGCGG	1804
Qy	310	ValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSer---	328
Db	1805	ATCAGCGCCGAAGGGCGCTCCACGACCGTGATCGCGCACCCGCTCTCGGTGAGCTTC	1864
Qy	329	-----ThrSerGlnLysAlaThrTyrSer	336
Db	1865	GGCGTCCAGCAGTGGCGCGCACACGACGAGCGCGGTTCACAAAGCAGCTGACCTACCG	1924
Qy	337	PheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp-----	353
Db	1925	AACCTGGCACCCAGGACGTCACGCTGNAGCTGACGTGACCGCCACCGACCCCAAGGGC	1984
Qy	354	-----AlaProAlaSer	357
Db	1985	AAGCGCGCCCGGGGCTTCTTCACGTGGCGCGCCACCGGTGACCGCTCCCGCGCGGC	2044
Qy	358	ThrThrAlaSerValThrLeuValAsnAspLeuAspLeu-----	370
Db	2045	GGCAGCGCTCCGTGACATGACCGCGACACCCGGCTCGCGCGACGCTGGACGGCGCG	2104
Qy	371	-----ValIleThrAlaProAsnGlyThrGlnTyrVal	381
Db	2105	TACTCGGCGTACGTTGTCGCGCACCGGGCGCGGACGCGTC	2146

APPLICANT: Akira ARISAWA et al.
 TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
 TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERI
 TITLE OF INVENTION: ITS EXPRESSION PRODUCT
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
 STREET: 2033 K Street, N.W. #800

QY 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGACCTGAAGGCGCGGTGACCGCGTCCAAAGAACTTACCGCGCGCCGCCG 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCCGGGACAAAGTGGGCCACGCAACCCAGTCGCTCGATCGCGCGGCGACGGCGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CATCTCAAGGGCAAGTACAAGGCGGTGCGACCGCGCGCGCGATCTCAACGGCAAGGTC 1231
QY 98 MetAspSerGlyGlyLeuGlyLysGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGACGACTCCGGT-----TTCCGGGACGACTCCGGCATCTCCCGGCATGGAGTGG 1285
QY 118 AlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTyrGlyAlaAlaValAsnGlyAl 137
Db 1286 GCGGCGCGGAGGCGCGGCGTGTCAATGATGCTGGCGGCGCATGGACACACCGCGAG 1345
QY 137 aTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLe 157
Db 1346 ACCGACCGGTGGAGGCGGGG-GTCGACAAGCTGTCCGCGGAGAGGGGCTCTGTTCGC 1404
QY 157 uPheAlaAlaGlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLys 177
Db 1405 CATCGCGCGCGCAACAGAGGCGCGGAG-----TCGATCGGTTCGCGCGGCGAGCGGA 1458
QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
Db 1459 GCGCGCCCTCACCGTCGCGGCC-----GTCCGA 1485
QY 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgI 216
Db 1486 CGACAAGGACAAGTCCGCCACTTCTCTCCACCGGCCCGCGCTCGGCGAGCGGCCCAT 1545
QY 216 eLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr 236
Db 1546 CAAGCCGGAGCTCACCGCTCCGCGGTGGACATCACGCGCGCTCGGCGGAGGGGCAACGA 1605
QY 236 oAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMe 256
Db 1606 CATCGCGCAGAGTCCGTGAGGAGCGCGCGGCTACATGACCATCTCCGCGACGTCGAT 1665
QY 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
Db 1666 GCGGACCCGCGACGTCGCGGCGCGCGGCGGCTCTCTGAAGCAGCAG-----1711
QY 276 gGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCCGACTGGACCTCCGCGGAACTGAAGGCGCGCTACCGGCTC 1758
QY 293 aAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAs 313
Db 1759 CACCAAG--GGCGGCAAGTACACCCGCTTCGAGCAGGGTTCGGGCGGATCCAGGCGCA 1815
QY 313 pLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSer-----328
Db 1816 CAAGGCGCTCAGCAGACCGTATGCGCGGACCGCGTCTCGGTGAGCTTCGGCGTCCAGCA 1875
QY 329 -----ThrSerGlnLysAlaThrTyrSerPheThrAlaTh 340
Db 1876 GTGGCGGCACACCGACGAGCGGTCAACCAAGCAGCTGACCTACCGCAACCTCGGCAC 1935
QY 340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp-----353
Db 1936 CCAGGAGCTCAGCTGAAGCTGAGTCGACGCGCACCGACCCCAAGGGCAAGGCGGCCCC 1995
QY 354 -----AlaProAlaSerThrAlaAla 361
Db 1996 GCGGGGCTTCTTACGCTGGGCGCCACCGGTGACCGTCCGCGGGCGGCGAGCGCCCTC 2055
QY 361 rValThrLeuValAsnAspLeuAspLeu-----Va 371

RESULT 15

US-09-514-340-1

; Sequence 1, Application US/09514340

; Patent No. 6361987

; GENERAL INFORMATION:

; APPLICANT: Akira ARISAWA et al.

; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC

; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE

; IIS EXPRESSION PRODUCT

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.

; STREET: 2033 K Street, N.W., #800

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20006

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Wordperfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/514,340

; FILING DATE: 28-Feb-2000

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/000,016

; FILING DATE: January 30, 1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Lee Cheng

; REGISTRATION NUMBER: 40,949

; REFERENCE/DOCKET NUMBER: <Unknown>

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-721-8200

; TELEFAX: 202-721-8250

; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2809 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: genomic DNA

; ORIGINAL SOURCE:

; ORGANISM: Streptomyces antibioticus

; STRAIN: <Unknown>

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 338..2539

; IDENTIFICATION METHOD: E

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 2540...2809

; IDENTIFICATION METHOD: P

; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-514-340-1

Alignment Scores:

Pred. No.: 5.9e-26

Score: 346.00

Percent Similarity: 42.00%

Best Local Similarity: 29.93%

Query Match: 15.40%

DB: 4

Length: 2809

Matches: 129

Conservative: 52

Mismatches: 167

Indels: 84

Gaps: 13

US-09-985-689A-1 (1-434) x US-09-514-340-1 (1-2809)

```
Qy 2 AspValAlaArgGlyLeuValAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACACGTCGTCGGGAGATCGCGCCCAAGGGGTGTGTCGCC---GGTACGACGGC 1009
Qy 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
Db 1010 AAGGGCGTGAAGATCGCGTCTGTGACACCGGTGTGACACGAGC----- 1054
Qy 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGGAGCTCAAGCGCGGGGTGACCGGCTCCAGAACTTCACCGCGCGCGCCGCG 1111
Qy 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCGGGGACAAAGTGGGCGCACCGCACCCACGTCGCTCGATCGCGCGGGGACGCGGCC 1171
Qy 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAAGGGCAAGTACAAAGGGCGTGCACCGCGCGCGATCTCAACGGCAAGGTC 1231
Qy 98 MetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGACGACTCGGT-----TTCGGCAGAGACTCCGCATCTCCCGCGCATGGAGTGG 1285
Qy 118 AlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
Db 1286 GCGGCGCGCAGCGCGCGACGTCGTCAACATGAGCTGGGCGGCATGACACACCGGAG 1345
Qy 137 aTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLe 157
Db 1346 ACGACCCCGCTGGAGCGCGC-GTCGACAAAGTGTCCGCGGAGAGGGCGCTCTGCTTCGC 1404
Qy 157 uPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys 177
Db 1405 CATCGCGCGCGCAACGAGGGCGCGGAG-----TCGATCGGTTCGCGCGAGCGCGGA 1458
Qy 177 ssnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
Db 1459 CGCGCGCCCTCACCGTCGGCGCC-----GTGCA 1485
Qy 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyValI 216
Db 1486 CGACAGGACAGCTCGCGGACTTCTCTCCACCGCGCCCGCTCGCGCGAGCGGCCAT 1545
Qy 216 eLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr 236
Db 1546 CAAGCGGACGTCACCGCTCCCGGCGTGACATCACGCGCGCTCGGCGAGGCGCAACGA 1605
Qy 236 oAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMe 256
Db 1606 CATCGCGCAGGAGTCCGTGAGGACCGCGCGCTACATGACCATCTCCGGCACGTGAT 1665
Qy 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
Db 1666 GCGGACCGCGCAGTCCGCGGCGCGCGCCCTCTGAAGCAGCAG----- 1711
Qy 276 gGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCCGACTGGACCTCCCGCACTGAAGGGCGCGCTCACCGGCTC 1758
Qy 293 aAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
Db 1759 CACCAAG---GGCGGCAAGTACACCCCGTTCGAGCAGGGTTCGGGCGGATCCAGGCCGA 1815
Qy 313 pLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSer----- 328
Db 1816 CAAGGGCTCCAGCAGACCGGTATCGCCAGCCCGGTCTCGGTGAGCTTCGGCGTCCAGCA 1875
Qy 329 -----ThrSerGlnLysAlaThrTyrSerPheThrAlaTh 340
Db 1876 GTGGCGGCACACCGACGAGCGGTGTCACCAAGCAGCTGACTTACCGCAACCTCGGCAC 1935
```

```
Qy 340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1936 CCAGGACGTCACGCTGAAGCTGACGTGACCGCCACCGACCCCAAGGCAAGCGCGCCC 1995
Qy 354 -----AlaProAlaSerThrThrAlaLe 361
Db 1996 GCGGGGCTTCTTCAAGCTGGCGCCACCAAGGTGACCGTCCCGCGGCGGCGGCGGCTC 2055
Qy 361 rValThrLeuValAsnAspLeuAspLeu-----Va 371
Db 2056 CGTCGACATGACCGCGGACACCGCGGTCTCGCGCGCACGGTGGACGGCGGTACTTCGGCGTA 2115
Qy 371 lIleThrAlaProAsnGlyThrGlnTyrVal 381
Db 2116 CGTGTGCTCCACCGGCGGCGGCGGCGGCGGTC 2146
```

Search completed: March 16, 2004, 01:18:00
Job time : 118 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 16, 2004, 01:15:59 ; Search time 359 seconds
(without alignments)
4450.734 Million cell updates/sec

Title: US-09-985-689a-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAINVPGVQTFSLAIVN 434

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Zgapop 6.0	Zgapext 7.0
Delop 6.0	Delext 7.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09985689/runat 10032004 112305 19840/app query.fasta_1.583
-DB=published Applications NA -QFMT=fastap -SUPFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=epct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pcp -NOR=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USERS=US09985689 @cgn 1 164 brunat 10032004 112305 19840
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARGE TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10:	/cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
17:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
18:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	2247	100.0	1305	15	US-10-385-662-1	Sequence 1, Appli
2	452.5	20.1	1977	13	US-10-090-624-11	Sequence 11, Appli
3	416.5	18.5	1236	13	US-10-090-624-2	Sequence 2, Appli
4	416.5	18.5	1982	13	US-10-090-624-15	Sequence 15, Appli
5	363.5	16.2	3624	14	US-10-156-761-5701	Sequence 5701, Ap
6	363.5	16.2	9025608	14	US-10-156-761-1	Sequence 3306, Ap
7	346.5	15.4	3417	14	US-10-156-761-3306	Sequence 1, Appli
8	346.5	15.4	9025608	14	US-10-156-761-1	Sequence 1, Appli
9	338	15.0	135638	14	US-10-314-657-1	Sequence 1334, Ap
10	332.5	14.8	1329	9	US-09-974-300-1934	Sequence 5, Appli
11	307	13.7	4765	13	US-10-090-624-5	Sequence 113, App
12	306.5	13.6	1560	15	US-10-084-848A-113	Sequence 2, Appli
13	306.5	13.6	59816	15	US-10-084-848A-2	Sequence 29, Appli
14	306.5	13.6	59816	15	US-10-084-848A-2	Sequence 33, Appli
15	283	12.6	3743	10	US-09-927-827-23	Sequence 1, Appli
16	282	12.6	3788	10	US-09-927-827-33	Sequence 5, Appli
17	276	12.3	1306	9	US-09-966-921A-1	Sequence 5384, Ap
18	276	12.3	1330	9	US-09-966-921A-5	Sequence 112429,
19	270	12.0	3303	14	US-10-156-761-5384	Sequence 1938, Ap
20	268	11.9	2192	12	US-10-424-599-112429	Sequence 1935, Ap
21	256	11.4	1485	9	US-09-974-300-1938	Sequence 1, Appli
22	253	11.3	1971	9	US-09-974-300-1935	Sequence 17, Appli
23	251	11.2	840	14	US-10-209-812-1	Sequence 34, Appli
24	248.5	11.1	2166	12	US-10-344-231-17	Sequence 11, Appli
25	248.5	11.1	2166	12	US-10-363-332A-17	Sequence 12, Appli
26	242.5	10.8	3884	10	US-09-927-827-34	Sequence 20, Appli
27	240	10.7	1140	8	US-08-322-678-11	Sequence 6, Appli
28	240	10.7	1140	8	US-08-322-678-12	Sequence 20, Appli
29	240	10.7	1140	16	US-10-323-324-11	Sequence 6, Appli
30	240	10.7	1140	16	US-10-323-324-12	Sequence 1, Appli
31	240	10.7	1143	14	US-10-313-853-6	Sequence 12, Appli
32	240	10.7	2588	12	US-10-344-231-20	Sequence 6, Appli
33	240	10.7	2588	12	US-10-363-332A-20	Sequence 20, Appli
34	237.5	10.6	1497	8	US-08-322-678-6	Sequence 6, Appli
35	237.5	10.6	1497	9	US-09-060-854B-1	Sequence 1, Appli
36	237.5	10.6	1497	14	US-10-033-325-1	Sequence 1, Appli
37	237.5	10.6	1497	14	US-10-228-572-1	Sequence 1, Appli
38	237.5	10.6	1497	15	US-10-423-649-1	Sequence 6, Appli
39	237.5	10.6	1497	16	US-10-323-324-6	Sequence 7, Appli
40	235.5	10.5	1149	15	US-10-146-908A-7	Sequence 1, Appli
41	235.5	10.5	1494	14	US-10-104-693-1	Sequence 10, Appli
42	232	10.3	3452	10	US-09-927-827-30	Sequence 5689, Ap
43	231.5	10.3	1332	14	US-10-156-761-5689	Sequence 9, Appli
44	230.5	10.3	1140	15	US-10-146-908A-9	Sequence 13, Appli
45	227.5	10.1	1140	9	US-09-920-118-13	

ALIGNMENTS

RESULT 1
US-10-385-662-1
; Sequence 1, Application US/10385662
; Publication No. US20040002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAKI, KATSUHIISA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938USO
; CURRENT APPLICATION NUMBER: US/10385,662
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231

```

; PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatenIn version 3.1
; SEQ ID NO 1
; LENGTH: 1305
; TYPE: DNA
; ORGANISM: Bacillus sp. KSM-KP43
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1305)
; OTHER INFORMATION:
US-10-385-662-1

Alignment Scores:
Pred. No.: 8,41e-242 Length: 1305
Score: 2947.00 Matches: 434
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-09-985-689A-1 (1-434) x US-10-385-662-1 (1-1305)

QY 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnSerSerTyGlyLeuTyr 20
D 1 AATGATGTTGCGGTGGAATTGTCAAAGCGGATGGCTCAGAGCAGCTACGGGTGTAT 60
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
D 61 GGCAAGGACAGATCGTAGCGTTCGGGATACAGGCTTGATACAGGTGCGCAATGACAGT 120
QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
D 121 TCGATCATGAAGCCCTTCGCGGGAAATTTCTGCATTATATGCAATATGCGACGGAGAT 180
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
D 181 AATGCCAATGATACGAATGGTTCATGTCAGCATGGTGGCTGGCTCGTATTAGGAACGCG 240
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
D 241 TCCACTAATAAAGGAATGCGCTCAGCGGAATCTAGTCTTCCCAATCTATCATGTGATAGC 300
QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
D 301 GGTGGGGACTTGAGGACTACCTTCGATCTGCAACCTTATTCAGCCNAGCATACAGT 360
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
D 361 GCTGGTGCCAGATTTCATACAACTCCTGGGAGCAGCAGTGAATGGGCTTACACAACA 420
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
D 421 GATTCAGAAATGTTGATGACTATGTCCGCAAAATGATGATGACGATCTTTTCGCTGCC 480
QY 161 GlyAsnGluGlyProAsnGlyGlyTrpIleSerAlaProGlyThrAlaLysAsnAlaIle 180
D 481 GCGAATGAAGACCGGACCGGACCACTCAGTGCACAGGACAGCAGCTAAAAATGCAATA 540
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
D 541 ACAGTCGGAGCTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATGCGGACAAATATCAAC 600
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
D 601 CATGTGCAAGTCTCTTTCACGTGGACCGACAAAGGATGGAGGATCAAAACCGGATGTC 660
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
D 661 ATGGACCGGGAACGTTCTATCATCAGCAAGATCTTCTTGCACCGGATTCCTCCCTTC 720
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260

```

```

RESULT 2
US-10-090-624-11
; Sequence 11, Application US/10090624
; Publication No. US2002013235A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090, 624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-11

Alignment Scores:
Pred. No.: 2,62e-40 Length: 1977
Score: 458.50 Matches: 138
Percent Similarity: 44.66% Conservatives: 67
Best Local Similarity: 30.07% Mismatches: 153
Query Match: 20.14% Indels: 101
DB: 13 Gaps: 18

```

US-09-985-689A-1 (1-434) x US-10-090-624-11 (1-1977)

QY 8 ValIyeAlaAsnValAlaGlnSerSeryrGlyLeuTyrrGlyGlnGlyGlnleValala 27
Db : : : : :
433 ATAGGGCCCATACCGTCTGGAACTCCCTCGGTACGACGAAGCGGTGGTTGCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArg 47
Db : : : : :
493 ATCGTCGATACGGGTATAGACGCAAC-----CACCCGATCTGAAG 534
QY 48 GlyIyelleThrAlaLeuTyrr--AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
Db : : : : :
535 GGCAGGTATAGCTGATAGCAGCCGCTCAACGCGCAGGTGACCCCCTACGATGACGAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
Db : : : : :
595 GGACACGGAACCAACGTTCCGGGTATCGTTGCCGAACCGGACGCGGTAACTCCAGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerlleMet-----AspSerGly 101
Db : : : : :
655 ATAGGGGTCCGCCCCCGCGCAAGTCGTCGCGGTCAAGGTTCTCGGTGCGCAGCGGTGCG 714
QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTy-SerAla 121
Db : : : : :
715 GGAAGGCTCCACCACATCATCGGGGTGTGACTGGTGCTGCCAGAACAGGACAGTAC 774
QY 122 GlyAlaArglle-----HisThrAsnSer 129
Db : : : : :
775 GGGATAGGGTATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGACGGAACCGACTCC 834
QY 130 TrpGlyAlaAlaValAlaAsnGlyAlaTyrrThrTrpAspSerArgAsnValAspTyrrVal 149
Db : : : : :
835 CTAGTCAGCGCCTCAACAGCCTGGGACGCC----- 867
QY 150 ArgLyAsnAspMetThrlleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr 169
Db : : : : :
868 -----GGTATAGTAGTCTCGTCGCGCGCGGCAACAGCGGGCCGCAACACCTACACC 918
QY 170 lleserAlaproglyThrAlalysAsnAlalleThrValGlyAlaThrGluaSnLeuArg 189
Db : : : : :
919 GTCGGCTCACCCCGCGCGGACGAGTCAATACGTCGTTGCA----- 963
QY 190 ProSerPheGlySerTyrrAlaAspAsnilleAsnHisValAlaGlnPheserSergly 209
Db : : : : :
964 -----GTTGACACAACGACACATCGCCAGCTTCTCCAGCAGGGA 1005
QY 210 ProThrlyAspGlyArgllelysProaspValMetAlaproGlyThrPheileuSer 229
Db : : : : :
1006 CCGACCGGACGGAAGGTCAACGCGAAGTCGTGCGCCCGCGGTGACATCATAGCC 1065
QY 230 AlaArgSerSerLeuAlaProaspSerSerPheTrpAlaAsnHisAspSerlystyrAla 249
Db : : : : :
1066 CCGCGCGCCAGC-----GGAACAGCATCGGCGACCCCGATAACGACTACTACACC 1116
QY 250 TyrMetGlyGlyThrSerMetAlaThrProlleValAlaGly---AsnValAlaGlnLeu 268
Db : : : : :
1117 AAGGCCCTCTGGAACACGATGGCCCGCAGCTTTCGGGGGTTTGGCGGCTCATCTCTC 1176
QY 269 ArgGluHispheVallylsAsnArgGlylleThrProLys-----ProSerLeuLeuLys 286
Db : : : : :
1177 CAGCCCCAC-----CCAGCTGACCCCGGACAGGTGAG 1212
QY 287 AlaAlaLeulleAlaGlyAla-----AlaAspIledglyLeu 298
Db : : : : :
1213 ACCGCCCTCATCGAGACCGCGCATAGTCGCCCCCAAGGATAGCGGACATCGCCTAC 1272
QY 299 GlyTyrrProasnGlnGlnGlyTyrrTpGlyArgValThrLeuAspLysSerLeu----- 316
Db : : : : :
1273 GGTGCG-----GGTAGGGTGAACGTCTACAGGCCCATCAAGTAC 1311
QY 317 ---AsnValAlaTyrrValAlaGlnUserSerSerSerSerSerGlnLysAlaThrTyrr 335
Db : : : : :
US-09-985-689A-1 (1-434) x US-10-090-624-2 (1-1236)
QY 12 ValAlaGlnSerSeryrGly-LeuTyrr-----GlyGlnGlyl 24
Db 24 GTCTGCAGCTCAAGTTATGCGCACTTACGTTTGGAACTTGGGATATGATGTTCTGGAAT 83
QY 24 nllleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisgl 44
Db : : : : :
84 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 125
QY 44 uAlaPheArgGlyLylleThrAlaLeuTyrrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db : : : : :
126 AGATCTCCAAGNAAGTA-----ATTGGTGGGTAGATTTTGTCAATGG 170
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValleuGl 78

```
Db 171 TAGGAGTTATCCATACGATGACCATGACATGGAATCATATAGCTTCAATAGCAGCTGG 230
Qy 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGGACGACCAAGTAATGCAAGTCAAGGGAATGGCTCCAGGAGCTAGCTGGCGGG 290
Qy 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyLeuProSerAsnLeuGl 112
Db 291 AATTAAAGGTTCTAGGTGCGGATGGTCTGGAAGCATATCTACTATATAATTAAGGAGTGA 350
Qy 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 351 GTGGCCCGTTGATACAAAGATAAGTACGGAATTAAGGTCTAATCTCTCTCTGTTCTC 410
Qy 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 411 AAGCCAGAGCTCAGATGCTAGTACGCTCTAAGTCAGGCTGTTAATGCAGGCTGGGATGC 470
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 471 T-----GGATTAGTTGTTGGTTGCCGC 494
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
Db 495 TGAACAACAGTGGACCTAACAGTATACAAATCGGTTCTCCAGCAGCTGCACCAAGTTAT 554
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACAGTTGGAGCC-----GTTGACAAGTATGA 581
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTATAACAAGCTTCTCAAGCAGAGGGCCAACTCCAGACGCGAGCTTAAGCTGAGGT 641
Qy 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 642 TGTGTCTCCAGGAATCGGATAATTGCTGCCAGACCAAGT-----GGAACCTAGCAT 692
Qy 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIl 260
Db 693 GGGTCAACCAATTAATGACTATTACACAGCAGCTCCTGGGACATCAATGGCAACTCTCA 752
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgLysHisPheValLysAsnArgIleThrPr 280
Db 753 CGTAGCTGTATGTCAGCCCTCTGCTCCAA-----GCACACCC 791
Qy 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 792 GAGCTGGACTCCAGACCAAAAGTAAACACAGCCCTCATAGAAACTGCTGATATCGTAAAGCC 851
Qy 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyValVa 310
Db 852 AGATGAAATAGCCGATATAGCTACGGTGCA-----GGTAGGGT 890
Qy 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
Db 891 TAATGCATCAAGGCTATAAAC-----TACGATAACTATGCAAAAGCTAGTGTTCACCTGG 944
Qy 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
Db 945 ATATGTTGCCAACAAAGGAGCAAACTCCACGTTCTGTTATTAGCGGAGCTTCGTTCTGT 1004
Qy 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1005 AACTGCCACATTATACCTGGGACAAATGCCAAT----- 1035
Qy 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
Db 1036 -AGCGACCTTGATCTTTACTCTTACCATCCCAATGGAAACCCAG---GTTGACTACTCTTA 1091
Qy 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
```

```
Db 1092 CACCGCTACTAT-----GGATTGAAAGGTTGGTTATTA 1127
Qy 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAACCCACTGATGGACATGACATTAAGTTTGTAGCTACAGC-----GGAAG 1178
Qy 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTATCAAGTAGATGTGTAAGT 1206

RESULT 4
US-10-090-624-15
; Sequence 15; Application US/10090624
; Publication No. US200203235A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-15

Alignment Scores:
Pred. No.: 2,82e-36 Length: 1962
Score: 416.50 Matches: 141
Percent Similarity: 42.55% Conservative: 59
Best Local Similarity: 30.00% Mismatches: 148
Query Match: 18.54% Indels: 122
DB: 13 Gaps: 20

US-09-985-689A-1 (1-434) x US-10-090-624-15 (1-1962)
Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyCl 24
Db 420 GTCTGCAGCTCAAGTTATGCGCACTTACGTTGGAACTTGGGATATGATGGTTCTGGAAT 479
Qy 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGl 44
Db 480 CACAAATAGGATTAATTGACACTGGAAATGAC-----GCTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAAAGGAAAAGTA-----ATTGGGTGGGTATGATTTTGTCAATCG 566
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db 567 TAGAGTTATCCATAGTACCATCGACATGGAACTCATGTAGCTTCAATAGCAGCTGG 626
Qy 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGGCAAGTACAAAGGAATGGCTCCAGGAGCTAAGCTGCGGG 686
Qy 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyGlyLeuProSerAsnLeuGl 112
Db 687 AATTAAAGTTCTAGTGGCCGATGTTCTTCTGGAAACCATATCTACTATAATTAAGGAGTTGA 746
Qy 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
```

```
Db 747 GTGGGCGCTGATACAAAGATAGTACGAATTAAGTTCATTAATCTTCTTGGTTC 806
Qy 132 a-
Db 807 AAGCCAGAGCTCAGATGCTAGCTGCTAAGTTCAGGCTGTTAATGACGCGTGGATGC 866
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnMetThrIleLeuPheAla 160
Db 867 T-
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 180
Db 891 TGGAAACAGTGGACCTTAACAGATACATCGCTCTCCAGCAGCTGCAAGCAAGTTAT 950
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 951 TACAGTTGGAGCC-
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyValArgLysLeuAspVa 220
Db 978 TGTTATACAAAGCTTCTCAGCAGAGGGCCAACTGCAGACGGGAGGCTTAAGCTGAGGT 1037
Qy 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTTGCTCCAGAACTGGAATAATTGCTGCAGAGCAAGT-
Qy 240 eTTPAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrPro 260
Db 1089 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCAACTCTCT 1148
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 1149 CGTAGCTGGTATTGCAGCCCTCTGCTCCAA-
Qy 280 oLys-
Db 1188 GAGCTGACTCCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247
Qy 294 -
Db 1248 AGATGAATAGCGCATATAGCTACGGTGCA-
Qy 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerLeuSerThrSe 330
Db 1287 TAATGCATACAAGCTATNAAC-
Qy 330 rGlnLysAla-
Db 1341 ATATGTTGCCAACAAAGCGACCAACTCACCAGTTCGTTATTAGCGGAGCTTCGTTG 1400
Qy 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATTATCTGGGACAAATGCCAAT-
Qy 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
Db 1432 -AGGACCTTGATCTTACCTTACGATCCCAATGGAAACAG-
Qy 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1488 CACCGCTACTAT-
Qy 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1524 CAACCCAACTGATGGAACTGGCAATTAAGTTGTAAGCTACAGC-
Qy 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1575 TGCAACTATCAAGTAGATGTGGTAACT 1602
RESULT 5
US-10-156-761-5701
```

```
; Sequence 5701, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5701
; LENGTH: 3624
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3624)
US-10-156-761-5701

Alignment Scores:
Pred. No.: 6,23e-30 Length: 3624
Score: 363.50 Matches: 128
Percent Similarity: 45.50% Conservative: 54
Best Local Similarity: 32.00% Mismatches: 145
Query Match: 16.18% Indels: 73
DB: 14 Gaps: 18

US-09-985-689A-1 (1-434) x US-10-156-761-5701 (1-3624)
Qy 8 VallysAlaAspValAlaGlnSerSer-----Tyr 17
Db 556 GTGAGGCGGACATGCGCGAGCAAGCGCAGATCGTACGCGGCGCGGTGGAGCGCC 615
Qy 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 616 GGGCTCACGGCGAGCGGTGCCTGCGCGTGTCTGACACCGCGGTGCGACACC----- 669
Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 670 -----ACTCACCCGACCTCGCGCGGCGGTGTCCCGAGCAAGAGCTTCATC 717
Qy 58 ArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 718 GACGGGGAGGAGTGCCTCGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGGCG 777
Qy 78 GlyAsnGlySer-----ThrAsnLysGlyMetAlaProGlnAlaAsnLeuVal 93
Db 778 GGCAGCGCGCGCTCCGACGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 837
Qy 94 PheGlnSerIleMet---AspSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGln 112
Db 838 CTCGCAAAAGTGCTCAGCAGCAGCGCGCGGGA-----AGCAGTCCCGAG 882
Qy 113 ThrLeuPheSerGlnAlaTyrSerAla-----GlyAlaArgIleHisThrAsnSer 129
Db 883 ATCATCGCGGCGATGGAATGGCGCGGAGCGGACGCGGCGGACGCGTCTCGATGAGC 942
Qy 130 TrpGlyAla-----AlaValAsnGlyAlaTyrThrAsp-----SerArgAsnVal 145
Db 943 CTCGGATCGACGAGCGGCGGCGGCGG-----ACCGACCCATGCGGCGGCGGCGG 993
Qy 146 AspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyPro 165
Db 994 GACACCTCTCGAGGAGACCGCGCGGCGCTCTTCTGTCGCGCGGGAACACCGGTC 1053
```


QY 285 LeuLeuAlaAala-----LeuLeuAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsn 302
 Db 6919614 ATGTCACGTCGAGCAACTCGACCTCCGTATATCATGTTGGGGGGGTGGGTGAGT 6919673
 QY 303 GlyAsnGlnGlyTyrGly---ArgValThrLeuAspIleSerLeuAsnValAlaTyr--- 320
 Db 6919674 GTGCCGAGCGGTGGCGCGGTACCGCGAGCGGCGACCGCGCTCCGCTCCAC 6919733
 QY 321 -----ValAsnGlnSerSerSerLeuLeuValAlaTyr--- 331
 Db 6919734 CGCTGGCCCATGACCGCATGACCGCTCGACGAGCGGTTCACCTACTCTCACTCC 6919793
 QY 332 LysAlaThrTyrSerPheThr---AlaThrAlaGlyLysProLeuLysIleSerLeuVal 350
 Db 6919794 GACACGACGCTGAGCTGAGCTCGCTGGGGCGCGCGCGGTGGTGGCCACCTC 6919853
 QY 351 TrpSerAspAla-----ProAlaSerThrThrAlaSerValThrLeuValAsnAsp 367
 Db 6919854 GCCGACCGCGCACTACCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGAC 6919913
 RESULT 7
 US-10-156-761-3306
 ; Sequence 3306, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 3306
 ; LENGTH: 3417
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(3417)
 US-10-156-761-3306
 Alignment Scores:
 Pred. No.: 4,6e-28 Length: 3417
 Score: 346.50 Matches: 137
 Percent Similarity: 43.81% Conservative: 54
 Best Local Similarity: 31.42% Mismatches: 177
 Query Match: 15.42% Indels: 70
 DB: 14 Gaps: 15
 US-09-985-689a-1 (1-434) x US-10-156-761-3306 (1-3417)
 QY 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
 Db 745 GGGTACGACGCGAGGGGCGTCAAGATCGCTGCGTGGACACCGGTGTCGAC----- 795
 QY 38 AsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
 Db 796 -----GCCACCCACCGGACCTCAAGACACCGAGTGGCGGAGTCCAGAACCTTCTCC 846
 QY 58 ArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
 Db 847 GCCCG 906
 QY 78 GlyAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 93

Db 907 GGCACCGCGCGCAAGTCCAAACGCGCAAGTACAAAGGTGTGCGCGCGCGCGCGAGTACCTC 966
 QY 94 PheGlnSerIleMetAspSerGlyGlyGlyLeuGlyGlyLeuProSerLeuLeuGlnThr 113
 Db 967 AACGCGAAGGTCTCGACGACACCGCGC-----TCGCGCGACGACTCCGCGCATCTGGCC 1020
 QY 114 LeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAl 133
 Db 1021 GGCATGAGTGGCGCGCGCGAGCGCGCGCGCGAGTCTGTCACCTGAGCTTGGCGCGCGC 1080
 QY 133 aValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAs 153
 Db 1081 GACACCCCGAGATCGACCGCTGGAAGCGAG-GTCAACAGCTCTCCGAGGAGAGGG 1139
 QY 153 pMetThrIleLeuPheAlaAlaGlyAsnGlyGlyProAsnGly---GlyThrIleSerAl 172
 Db 1140 CATCTCTTCGCGATCGCGCGCGCGCAAGCGGAGTTCGGCGAGAGACCTATCGGCTC 1199
 QY 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh 192
 Db 1200 CCGCGGACGCGCGCGCGCTCACCGTCCGCGCGC----- 1236
 QY 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrIly 212
 Db 1237 -----GTGAACGACGACGAGCAAGCTGGGTCTCTTCACGCGCGCGCGCGCTC 1286
 QY 212 aAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSe 232
 Db 1287 GGACGCGCGCATCAAGCGCGCGCTCACCGCACCGCGGTGGAGATCACCGCGCGC----- 1341
 QY 232 rSerLeuAlaProAspSerSerPheTyrAlaAsnHisAspSerLys-----TyrAl 249
 Db 1342 -----GCGCGCGCGCGCGCGCTCATCGACGAGGAGTGGCGAGCGCGCGCGCTACCT 1397
 QY 249 aTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuAr 269
 Db 1398 CACCATCTCCGCTACGTGCGTGGCGACCGCGCATGTGCGCGCGCGCGCGCGCTCTCAA 1457
 QY 269 gGluHisPheValLysAsnArgIleThrProLysProSerLeu-----LeuLysAl 287
 Db 1458 CGACGACGAC-----CCCACTGGTCTTTCGCGGAGTCAAGGG 1496
 QY 287 aAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTr 307
 Db 1497 CGCGCTGACCGCGCTCCCGAAG-----GGCGGCAAGTACACCGCGTTCAGCAGCGCTC 1550
 QY 307 pGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlnSerSerLe 327
 Db 1551 GGGCGGTATCGCGTGCACAGGCGATCAAGCAGTCCGTGATCGCCACCGCGCGCTCGGT 1510
 QY 327 uSerThrSerGlnLysAlaThrTyrSerPhe-ThrAlaThrAlaGlyLysProLeuLysI 347
 Db 1611 G-----AGCTTCGGCATCCAGCAGTGGCGCGCACACCGACGA 1646
 QY 347 leSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuValAsnA 367
 Db 1647 CAAGCGGTCAACCGCAGCTACCTACCGCAACCTCGCGACGAGTGACGTC-ACGCTGA 1705
 QY 367 spLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPhe---- 385
 Db 1706 ACCTCGCGTCCAGCGCGACCAACCCCAAGGGCGTGGCGGTCCGCTCGGGCTTCTTCAAGC 1765
 QY 386 -----ThrSerProTyrAsnAspAsnTyrAspGlyArgAsnValG 400
 Db 1766 TCGGCGCGACGAGGTACGGTCCCGCGCGC-----GGCAAGGCTTCGGTCG 1813
 QY 400 lu-----AsnValPheIleAsnAlaProGlnSerGlyThrThrIleGluValG 417
 Db 1814 ACTTCAGCGTCAACACGAGCTGGCGCGCACCGACGCGCGGTACTCGCGCTGCTGTA 1873
 QY 417 lnAlaTyrAsnValproValGlyProGlnThrPheSerLeuAla 431


```
; PRIOR APPLICATION NUMBER: US 60/278,935
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 1
; LENGTH: 135638
; TYPE: DNA
; ORGANISM: Streptomyces atroolivaceus
US-10-314-657-1

Alignment Scores:
Pred. No.: 1,05e-24 Length: 135638
Score: 338.00 Matches: 127
Percent Similarity: 43.06% Conservative: 59
Best Local Similarity: 29.40% Mismatches: 144
Query Match: 15.04% Indels: 102
DB: 17 Gaps: 17

US-09-985-689A-1 (1-434) x US-10-314-657-1 (1-135638)

QY 6 GlyIleValIysAlaAspValAlaGlnSer----- 15
Db 8067 GGCAGGGTGAAGCCGATCTGGCCGACTCCACCGCCAGATCGGCGCGCAGAGGTATGG 8126

QY 16 SerTyGlyLeuTyrglyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35
Db 8127 GCGAGGGCCACACCGCCAGGACGTGAAGTCCGATCTGCACAGCGCGCGGACACC 8186

QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIle-----ThrAlaLeu 53
Db 8187 -----GAAACCCCGACCTGCTGGGCGAGGTGCCAGACGCCAGC 8228

QY 54 TyrAlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAla 73
Db 8229 TTCGTCCCGCGGAGGACGACATCGCC---GACTACAACGCCACGCCAGCAGCTCGCC 8285

QY 74 GlySerValLeuGlyAsnGlySerThrAsp-----LysGlyMetAlaProGln 89
Db 8286 TCACCATCTGCGCAGCGGCGGCTCCGACGGCAGAGCGGGGTGTCGCTCCGCG 8345

QY 90 AlaAsnLeuValPheGlnSerIleMetAspSerGlyGlyGly----- 103
Db 8346 GCCCGGTGTCGTCGCGGAGGTGCTCACTCCAGGGCAGCGGCCAGGAATCGTGGATC 8405

QY 104 LeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTySerAlaGlyAla 123
Db 8406 ATCCCGGCGCATGGAG-----TGGCCCGCCCGCGCAGCAGCAAGAGGCC 8444

QY 124 ArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyThrThrAsp----- 141
Db 8445 AGGATCATCAGCATGAGCCTGGGC-----GGCGCGGTGACAAAGACCGCCGATG 8495

QY 142 SerArgAsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 161
Db 8496 AGCCAGGCGCTCGACGAACTCAGCCACGACCGCGGCGGTGTCGTGTGATCGCGCGGGC 8555

QY 162 AsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 181
Db 8556 AACCGCGCCCGCAC-----TCCATCAGCAGCCCGGTGCGGAGACTCCGCGCTGACC 8609

QY 182 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 201
Db 8610 GTCGGCGCC-----GTCGACTCCACCGCACAGC 8636

QY 202 ValAlaGlnPheSerSerArgGlyProThrLys---AspGlyArgIleLysProAspVal 220
Db 8637 CTCGCGCACTTCTCCAGCGGGCCCGGTGACGGCGACCGCGGGCTGAAGCCGGAGATC 8696

QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 8697 ACCGCGCGCGCTCGCATCGTCGCGCGCGCTCGCACTACAAGCGCGGCTCCGCGTAC 8756

QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyMetGlyThrSerMetAlaThrProIle 260
Db 8757 -----TACACCACTGAGCGGCACGCTCGATGGCGCGCGCAC 8795

QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyIleThrPro 280
Db 8796 GTCGCCGTGTGCGCGGCTCTCGCCCGCAGACCCCGACTGGACGGGACCCAGCTC 8855

QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 8856 AAGGAGGCACTGGTCAGCAGCGCCAGGCAAGCGCGCG-----TAC 8897

QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTy 320
Db 8898 ACCCGGTACAGCGCGCGCGCGCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 8957

QY 321 ValAsnGluSerSerSerLeuSer-----ThrSer 330
Db 8958 TTCGCCACCGACCGCGCTACTCCGGTTCACACGTGGCCCGCGAAGCCCGGGGAGACC 9017

QY 331 GlnLysAlaThrTyrSerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuVal 350
Db 9018 GATGTCGGAGCGGTGACATCACCAACGTGCGGACGCCCGCGTCAACCTCGCC 9077

QY 351 TrpSer-----AspAlaPro 355
Db 9078 GTCACCGCACCGTCCCGCGGGTGTTCAGCTCTCCGAGGACCATGTACCGTGGCC 9137

QY 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAsp----- 369
Db 9138 GCGCACGGCAGCGCGCGGTCCACCTGACCGCGGCTCTGGACACAGCTGGCGGGCGAC 9197

QY 370 -----LeuValIleThrAlaProAsnGlyThr 378
Db 9198 TCGGTACGCGCGGTGATCACCGGTACGGACGGCGACG 9233

RESULT 10
US-09-974-300-1934
; Sequence 1934, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berk, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1934
; LENGTH: 1329
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1934

Alignment Scores:
Pred. No.: 4.12e-27 Length: 1329
Score: 332.50 Matches: 104
Percent Similarity: 43.53% Conservative: 44
Best Local Similarity: 30.59% Mismatches: 109
Query Match: 14.80% Indels: 83
DB: 9 Gaps: 13

US-09-985-689A-1 (1-434) x US-09-974-300-1934 (1-1329)

QY 11 AspValAlaGlnSerSerTyGlyLeuTyrglyGlnIleValAlaValAlaAsp 30
Db 406 GAAGTGGTCAGAAACATCAGACGCTGACAGCAAGAGTGACAGTGGCTGCTATTGAT 465
```

```

QY 31 ThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIle 50
Db 466 AGCGCGGTA-----TACCCCTCAGAGATCTTGAAGCGAGATC 504
QY 51 ThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAspThrAsnGlyHisGly 69
Db 505 AGGGCTTTCAAGACTTTATCACCAGAGACAGAACCCCTATGATGACATGGCGACGGC 564
QY 70 ThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsn-----LysGly 85
Db 565 ACACACTGCGCGGTGATGCTTGGGAAACAGGAGCGGCTCATCGGGTCAGTACCGCGGA 624
QY 86 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGlyGlyLeuGly 105
Db 625 CCTGCTCCAGACGAACTTGCTGGTAAAGATTGGACAAATGGGATCC---GGA 681
QY 106 GlyLeuProSerAsnLeuGln-----112
Db 682 TCGCTCGAAACCGTCATTCAGGCGTAGATTGGTGCATTCATTCATTAAGGAAATCCT 741
QY 113 -----ThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsn 128
Db 742 GATGATCCGATCGACATTATTTCAATGATTTGGGTGCGAAGCCTTCCGCTACGAGAT 801
QY 129 -----SerTrpGlyAlaAlaValAsnGly 136
Db 802 GAAGAAGAGATCCAGTCGTAAAGCTGTTTCATGCGAGCATGGACCGCAGGCATT----- 855
QY 137 AlaTyrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIle 156
Db 856 -----GTTGTA 861
QY 157 LeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAla 176
Db 862 TGTGCGGCGACCGCACTCCGGTCTGATGCGCAACGATTCGCGCCGGGTGTCAGC 921
QY 177 LysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAla 196
Db 922 AGCAGATTATTACAGTCGGAGCCTTGGATGAC---AGGGATACAGTCAGCGCGGAGGAT 978
QY 197 AspAsnIleAsnHisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIle 216
Db 979 GAGCAT-----GTGCGCTTATTCAACGAGAGGCGCCACCAATCTATGTTCAATC 1029
QY 217 LysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer-----SerLeu 234
Db 1030 AAACCGGACTTGCTGTGACCGGCGCAAAATATTACGTCGCTTCGTTACCGCGATCTTT 1089
QY 235 AlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThr 254
Db 1090 CTCGATAAGCTGCAGAAACAAACAGAGTCGGCAGCAAAATATATGACATGTCCGGAACC 1149
QY 255 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLys 274
Db 1150 TCGATGGCTACGCGGATCTGCGCAGGAATTCGCGCA-----CTTATCCTT 1194
QY 275 AsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAla 294
Db 1195 CAGCAAGCCCGCGCACAGAACCTGTATGAGTCAAAACACCTGCTATATGAGCGGT---ACC 1251
QY 295 AspIle-----GlyLeuGlyTyrProAsnGlyAsnGln 305
Db 1252 GATTATGGAAGATCGCGATCCAAATGTTTACGGTGGCGGATACATCAACGACAGAACAA 1311

```

RESULT 11

```

US-10-090-624-5
; Sequence 5, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo

```

```

; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 4765
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; US-10-090-624-5

Alignment Scores:
Pred. No.: 2,04e-23 Length: 4765
Score: 307.00 Matches: 142
Percent Similarity: 38.92% Conservative: 60
Best Local Similarity: 27.36% Mismatches: 178
Query Match: 13.66% Indels: 139
DB: 13 Gaps: 21

US-09-985-689A-1 (1-434) x US-10-090-624-5 (1-4765)

QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1024 GGCAATGGTTATGACATTGTCATATGCGATCTGCTTGTACCTGACCTTACCAGCTTACCAGCGAA 1083
QY 41 -----SerMetHisGluAlaPheArgGly 48
Db 1084 GTTCCACTTGGCGCAGTACACGTTACTTATGATGTTGCTGTTTGTACTACTACTACGCT 1143
QY 49 LysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-----63
Db 1144 CCTCTCAACTACGCTGTTCGAGAAATAGATCTTACGGAGATATCCAGTATTTGGGTGG 1203
QY 64 AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsn 83
Db 1204 GATGGTCACGGTCACGGAACTCAGTAGCTGGAACTGTTGCTGGTTACGACGACCAAT 1263
QY 83 -----83
Db 1264 GATGCTTGGGATGGCTCAGTATGTACTCTGGTGAATGGGAAGTGTCTCAAGACTCTAT 1323
QY 84 -----LysGlyMetAlaProGlnAlaAsn 91
Db 1324 GGTGGGATTTACGAACGTTTACCAGACACCGCTGCGAGGTGTTGCTCCAGGTGCCAA 1383
QY 92 LeuValPheGlnSerIleMetAspSerGlyGlyLeuGlyGlyLeuProSerAsnLeu 111
Db 1384 ATATGGCAATAAGATTCTTAGGAGT---GATGGACGGGTAGCTAGTGGGATATATATA 1440
QY 112 GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGly 131
Db 1441 GAAGGTATG---ACATACGACGCAACCCATGGTGGCAGACGTTATAGCATGCTCTCGT 1497
QY 132 AlaAlaValAsnGlyAlaTyrThr-----ThrAspSerArgAsn-----ValAspAsp 147
Db 1498 GGA-----AATGCTCCATCTTAGATGGTACTGATCCAGAAAGCGGTGCTGTGGATGAG 1551
QY 148 TyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGly 167
Db 1552 CTTACCGAAAGTACGGTGTGTATTCGTATAGTTCAGAGAAATAGAGTCTCTGGCATT 1611
QY 168 GlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsn 187
Db 1612 AACATCGTTGGAAGTCTCTGGTGTGGCAACAAAGCAATAACTGTTGGAGCTGTGCA---1668

```

```
QY 188 LeuArgProSerPheGlySerTyrAlaAsp----- 197
Db 1669 GTGCCAATTAAAGTTGAGTTATTTCCCAAGCACTTGGATATCTCGATTACTATGGA 1728
QY 198 -----AsnIleAsnHieValAlaGlnPheSerSerArgGly 209
Db 1729 TTCATTACTTCCCGCTACACAAAGTT--AGAATAGCATCTTCTCAAGCAGAGGG 1785
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
Db 1786 CCGAGAATAGATGGTGAATAAACAACCAATAGTAGTGGCTCCAGGTTACGGAATTACTCA 1845
QY 230 AlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHieAspSerLysTyrAla 249
Db 1846 TCCTCGCGATGAGTTGGCGGAGCTGACTTC----- 1878
QY 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArg 269
Db 1879 ---ATGCTGGAACCTCGATGGCTACTCCACATGTCCAGCGGTGCTCGTCCATCTCTCAT 1935
QY 270 GluHisPheValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu 289
Db 1936 AGCGGG---GCAAGCGCGAGGGAATATACTACATCCAGATATATTAAGAGGTTCTT 1992
QY 290 IleAlaGlyAlaAlaAspIle-----GlyLeuGlyTyrProAsnGly 303
Db 1993 GAGAGCGGTGCAACCTGGCTTGAGGAGATCCATATACTGGCGAGAGTACACTGAGCTT 2052
QY 304 AsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlu 323
Db 2053 GACCAAGGTCATGGTCTGTGAACGTACCAAGTCTCGGAATCTTAAAGCTATATAAC 2112
QY 324 SerSerSerLeuSerThrSerGln-LysAlaThrTyrSerPheThrAlaThrAlaGlyLys 343
Db 2113 GGCACCACTCTCCAAATGTTGATCTACTGGCGACAGACAGTCTTACAGCGACT-- 2164
QY 343 sProLeuLysIleSerLeuValThrSerAspAlaProAlaSerThrAlaSerValTh 363
Db 2165 ---TTGCGGAGTACTGGGTGTGGAGCTTATAAGAGGTCTCTACGCAAGGAACT-- 2215
QY 363 rLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThr--GlnTyrValGly 382
Db 2216 -----CTATACCTGACATTTTGAGATCTATGCAACTGAGCCATGGATTAAAGCCTTTTGT 2316
QY 383 Asn-----AspPheThrSerProTyrAsnAspAsnTrp----- 393
Db 2257 GACACGGAGTACAGAACTTTTGAGATCTATGCAACTGAGCCATGGATTAAAGCCTTTTGT 2316
QY 394 AspGly-----ArgAsnAsnValGluAsnValPhe-----Ile 404
Db 2317 AGTGAAGTGTAAATTCAGAGAACATACCAGATTGTCTTAGGTGAATATGATGTA 2376
QY 405 AsnAlaProGlnSerGlyThrTyr----- 412
Db 2377 GAGGCTCTGAGCGAGTCTCTATGTTGAAGGATAATCATTTGATGATCCAAACAGCCA 2436
QY 413 ThrIleGluValGlnAlaTyrAsnValProValGlyProGlnThrPheSer 429
Db 2437 GTTATTGAAGACGAGATCTTGAACACAAATTGTTATTCGCGAAGTTCACT 2487

RESULT 12
US-10-084-846A-113
; Sequence 113, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; CURRENT FILING DATE: 2003-02-25
```

```
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 113
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: avix16 dna: partial sequence of coding strand 1; nucleotide 1
; OTHER INFORMATION: corresponds to nucleotide 56,198 of coding strand 1.
US-10-084-846A-113

Alignment Scores:
Pred. No.: 4,32e-24 Length: 1560
Score: 306.50 Matches: 116
Percent Similarity: 39.20% Conservative: 51
Best Local Similarity: 27.23% Mismatches: 154
Query Match: 13.64% Indels: 105
DB: 15 Gaps: 19

US-09-985-689A-1 (1-434) x US-10-084-846A-113 (1-1560)
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 451 GGGCAGGAGTACGGCGGTACTGTCATCGACACACCGCGCTC-----CGC 492
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 493 ATCACCCACAGCGACTTTCGGCGCGCGCTCTCCGCTACGCGCATCGACGCGACGAC 552
QY 60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 553 AACACCGCCAGGACGCGCACCGCGCACGCGCACGCGCGCGCACGCGCGCGCAAC 612
QY 80 GlySerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 99
Db 613 GCC-----TACGGCGTCCCAAGAGCGCAAGATCGTAGCGCTCCGGTCTCGAAC 663
QY 100 SerGlyGly-----GlyLeuGlyGlyLeuProSerAsn 110
Db 664 AACTCCGCGCAGGCGACACCGCCCGCTGTCGCCGCGCATCGTGGTCCGCGCAAC 723
QY 111 LeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrp 130
Db 724 -----GCCGTCAAGCGCGCGCTGCGCAACATGTCCTCCTC 756
QY 131 GlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArg 150
Db 757 GCGCGCGCGCGCGACCGCCCTCGACACGCGCGCTACGCAAC-----GCCATG 804
QY 151 LysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIle 170
Db 805 GCCTCCGCGCTCACCTTCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 864
QY 171 SerAlaProGlyThrAlaAlaAsnAlaIleThrValGlyAlaThr-----GluAsnLeu 188
Db 865 TCA-----CCCGCAGCGGTCCAGCGCCATACGTTGCGCGCGCGCGCGCGCGCGCG 921
QY 189 ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
Db 922 AAGCGCGGTACTCTCAACTACGCTCGCTCCTC----- 954
QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeu 228
Db 955 -----GACCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 984
QY 229 SerAlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHieAspSerLysTyr 248
Db 985 TCGGCC-----TGGAACCTCAAGCGACTCGCGCGAC 1014
```

QY 249 AlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 268
 DB 1015 AACACCATCTCCGATACGTGCGACCGCGGCGGCGGCGGCTC--- 1071
 QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeu 285
 DB 1072 -----CACCTCGCGCCCAACCCCTCGGCCACCCCGCTCCAGGTCGCGGCGCTGACG 1125
 QY 286 LysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGln 305
 DB 1126 TCCGCGCCACACCGCGGTGTCTACCAACCCCGCGGCGGCTCGCCAC- 1176
 QY 306 GlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaIleValAsnGlu----- 323
 DB 1177 -----CGGCTCTGTACTACGTGCGCGCGCGGCAC 1203
 QY 324 -----SerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAla 339
 DB 1204 GACACCTCCGCGCGCGCTTCGAGAACACCGGTGACTACAGTACGCGCAACTCC 1263
 QY 340 ThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThr 359
 DB 1264 ACGTCGAGTCCCGGTACGCTCCGCGCTCCGCGCAACGCGCGCTCGCCCTCGCC 1323
 QY 360 AlaSerVal-----ThrLeuValAsnAspLeuLeuValIleThrAlaPro 375
 DB 1324 GTAGAGGTCCACATCGTCCACCGTACGCGGACCTCCAGGTCACGCTGATCGCCCCC 1383
 QY 376 AsnGlyThrGlnTyrValGlyAsnAspPhe---ThrSerProTyrAsnAspAsnTrpAsp 394
 DB 1384 GACGCGACGCGGTACACGCTCAAGTCGATCGCGCACCGCGCGCGAGTTCGGAC- 1434
 QY 395 GlyArgAsnAsnValGluAsnValPhe---IleAsnAlaProGln-----SerGly 410
 DB 1435 -----ACATCAACACACACGACTCGGTGAAACGCTCTCTCGAGGCGCGCAACGCG 1485
 QY 411 ThrTyrThrIleGluVal 416
 DB 1486 ACGTGGAACTCGGGTG 1503

RESULT 13

US-10-084-846A-1

; Sequence 1, Application US/10084846A

; Publication No. US20040006026A1

; GENERAL INFORMATION:

; APPLICANT: WEITNAUER, GABRIELE

; APPLICANT: TREFFER, AXEL

; APPLICANT: MUHLWEG, AGNES

; APPLICANT: BECHTOLD, ANDREAS

; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES

; FILE REFERENCE: 1974-005

; CURRENT APPLICATION NUMBER: US/10/084,846A

; CURRENT FILING DATE: 2003-02-25

; PRIOR APPLICATION NUMBER: PCT/EP01/09815

; PRIOR FILING DATE: 2001-08-24

; PRIOR APPLICATION NUMBER: DE 101 09 166.4

; PRIOR FILING DATE: 2001-02-25

; NUMBER OF SEQ ID NOS: 120

; SOFTWARE: Patentin Ver. 3.2

; SEQ ID NO 1

; LENGTH: 59816

; TYPE: DNA

; ORGANISM: Streptomyces viridochromogenes

US-10-084-846A-1

Alignment Scores:

Pred. No.: 1,04E-21 Length: 59816

Score: 306.50 Matches: 116

Percent Similarity: 39.20% Conservative: 51

Best Local Similarity: 27.23% Mismatches: 154

Query Match: 13.64% Indels: 105

DB: 15 Gaps: 19

US-09-985-689A-1 (1-434) x US-10-084-846A-1 (1-59816)
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 56648 GGGGAGGAGTACGCGGTACGTATCATCACACCGGCGTC-----CGC 56689
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
 DB 56690 ATCACCCACAGCGACTTCGGCGCGCGGCTCTCTACGCGTACGAGCCATCGAACACGAC 56749
 QY 60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
 DB 56750 AACACCGCCACGACGCGCCACGCGCACGCGACGCGTGGCGCGGACCGGTCCGCGGCAAC 56809
 QY 80 GlySerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 99
 DB 56810 GCC-----TACGGCGTCCCAAGAAAGCAAGATCGTAGGGGTCCGGGTGTGTGAAC 56860
 QY 100 SerGlyGly-----GlyLeuGlyGlyLeuProSerAsn 110
 DB 56861 AACTCCGGCCAGGCGCACACCCCGCGGTCTGCGCGGATCGACTGGGTCCCGCGGAC 56920
 QY 111 LeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrp 130
 DB 56921 -----GCCGTCAAGCGCGCTCGCCCAACATGTCTCCCTC 56953
 QY 131 GlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArg 150
 DB 56954 GCGCGCGCGCGCACGCGCCCTCGACACGCGCGTACGCAAC-----GCCATG 57001
 QY 151 LysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIle 170
 DB 57002 GCCTCCGGCTCACCTTCGCGTGGCGCGCGCAACGAGTCGACCAACGCTCCACGAGG 57061
 QY 171 SerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeu 188
 DB 57062 TCA---CCGCGACGCGTCCACGAGGCCATCGGTCGCGCGCGCACCAACGCTCGGACGCGC 57118
 QY 189 ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
 DB 57119 AAGCGCGGTACTCCAACTACGCTCGGTCTC----- 57151
 QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeu 228
 DB 57152 -----GACCTTCGCGCGCGGTTCGCTCCATCAC 57181
 QY 229 SerAlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyr 248
 DB 57182 TCGGCC-----TGAACCTCAAGCGACTCGGCGAC 57211
 QY 249 AlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 268
 DB 57212 AACACCATCTCCGCTACGTGCGACCGCGCGCGCGCGCGCGCGCTC--- 57268
 QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeu 285
 DB 57269 -----CACCTCCCGCCCAACCCCTCGGCCACCCCGTCCCGAGGTCCCGCGGCTGACG 57322
 QY 286 LysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGln 305
 DB 57323 TCCGCGCGCACACCGCGGTGTCTACCAACCCCGCGCGCGGTTCGCCCAAC----- 57373
 QY 306 GlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlu----- 323
 DB 57374 -----CGGCTCTGTACTCGCGCGCGGCGAC 57400
 QY 324 -----SerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAla 339
 DB 57401 GACACCCCTCCGCGCGCGCTTCGAGAACACCGGTGACTACACGATCAGCAGCAACTCC 57460
 QY 340 ThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThr 359
 DB 57461 ACGGTGAGTCCCGGTGACGCTCTCGCGGTCTCCGCGCAACGCGCGCTCGGCGCTCGCC 57520

QY 360 AlaSerVal-----ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
 Db 57521 GTAGAGTCCACATCGTCCACACGTATACGCGGACCTCCAGTCCAGCTGATCGCCCC 57580
 QY 376 AsnGlyThrClnTyValGlyAsnAspPhe---ThrSerProTyAsnAspAsnTrpAsp 394
 Db 57581 GACGCGACGGCGTACAGCTCAAGTCGTACGCGACCGCGGCGGAGTTCGGAC----- 57631
 QY 395 GlyArgAsnAsnValGluAsnValPhe---IleAsnAlaProGln-----SerGly 410
 Db 57632 -----AACATCAACACCGTACTCGGTGAACGCTCTCTCGGAGCGGCCAACGGC 57682
 QY 411 ThrTyThrIleGluVal 416
 Db 57683 ACGTGGAACCTGCGGGTG 57700

RESULT 14
 US-10-084-846A-2/c
 ; Sequence 2, Application US/10084846A
 ; Publication No. US20040006026A1
 ; GENERAL INFORMATION:
 ; APPLICANT: WEITNAUER, GABRIELE
 ; APPLICANT: MUHLENWEG, AGNES
 ; APPLICANT: TREFFZER, AXEL
 ; APPLICANT: BRECHTHOLD, ANDREAS
 ; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
 ; FILE REFERENCE: 1974-005
 ; CURRENT APPLICATION NUMBER: US/10/084,846A
 ; CURRENT FILING DATE: 2003-02-25
 ; PRIOR APPLICATION NUMBER: PCT/EP01/09815
 ; PRIOR FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: DE 101 09 166.4
 ; PRIOR FILING DATE: 2001-02-25
 ; NUMBER OF SEQ ID NOS: 120
 ; SOFTWARE: Patentin Ver. 3.2
 ; SEQ ID NO 2
 ; LENGTH: 59816
 ; TYPE: DNA
 ; ORGANISM: Streptomyces viridochromogenes
 US-10-084-846A-2

Alignment Scores:
 Pred. No.: 1.04e-21 Length: 59816
 Score: 306.50 Matches: 116
 Percent Similarity: 39.20% Conservative: 51
 Best Local Similarity: 27.23% Mismatches: 154
 Query Match: 13.64% Indels: 105
 DB: 15 Gaps: 19

US-09-985-689A-1 (1-434) x US-10-084-846A-2 (1-59816)

QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 3169 GGGCAGGAGTGAGCGGCTACGTACATCGACACCGCGTC-----CGC 3128
 QY 41 SerMethHisGluAlaPheArgGlyIleThrAlaLeuTyR---AlaLeuGlyArgThr 59
 Db 3127 ATCACCCACACGCGATTTCGCGCGCGCGCTCTCTACGCTTACGACGCCATCGACAACGAC 3068
 QY 60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
 Db 3067 AACACCGCCAGAGCGCCACGCGGACGCGGACGCGTGGCGGCGGACGCTCGCGCGCAC 3008
 QY 80 GlySerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 99
 Db 3007 GCC-----TACGCGCGTCCGACAGAGCCCAAGATCGTAGCGCTCGCGGTGTAAC 2957
 QY 100 SerGlyGly-----GlyLeuGlyGlyLeuProSerAsn 110
 Db 2956 AACTCCGCGCAGGCACACCGCCAGGTGCTCGCGGCATCGACTGGGTGCGCCCGGAAAC 2897
 QY 111 LeuGlnThrLeuPheSerGlnAlaTyRSerAlaGlyAlaArgIleHisThrAsnSerTrp 130

Db 2996 -----GCCGTCAAGCGCGCGTCCGCCAACATGTCCTC 2864
 QY 131 GlyAlaAlaValAsnGlyAlaTyThrThrAspSerArgAsnValAspAspTyRValArg 150
 Db 2863 GCGCGCGCGCGACACGCGCTCGACACGCGGTACGCAAC-----GCCATG 2816
 QY 151 LysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIle 170
 Db 2815 GCCTCCGCGCGTCACTTCGCGCGCGCGGCAACAGTCCGACCAACGCTCCACGAGG 2756
 QY 171 SerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeu 188
 Db 2755 TCA---CCCGACCGCGTCACCGAGGCCATCACGTCGCGCGACGACGAGTCGAGCGC 2699
 QY 189 ArgProSerPheGlySerTyRAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
 Db 2698 AAGCGCGGCTACTCCAACTACGGTCCGTCTCTC----- 2666
 QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeu 228
 Db 2665 -----GACCTCTCCCGCGCGGTTCCTCATCACC 2636
 QY 229 SerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyR 248
 Db 2635 TCGGCC-----TGGAACTCAAGCAGCTCGCGCACC 2606
 QY 249 AlaTyMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 268
 Db 2605 AACACCATCTCCGTACGTGATGCGACCCCGCAGCTGGCGGCGCGCGCTC--- 2549
 QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeu 285
 Db 2548 -----CACCTCGCGCGCAACCCCTCGCGCACCCCGTCCAGGTCCGACGCGCTCAGC 2495
 QY 286 LysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyRProAsnGlyAsnGln 305
 Db 2494 TCGCGCGCCACACCGCGCTCGTCACCAACCCCGCGCGCTCGCCCAAC----- 2444
 QY 306 GlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyRValAsnGlu----- 323
 Db 2443 -----CGGCTCTGTACGTCCGCGCGCGCAC 2417
 QY 324 -----SerSerSerLeuSerThrSerGlnLysAlaThrTyRSerPheThrAla 339
 Db 2416 GACCACCTCCGCGCGCGCTTCGAGAACACACCGGTACGTACACGATCAACGCAACTCC 2357
 QY 340 ThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThr 359
 Db 2356 ACGGTGAGTCCCGGTGACGCTCTCGCGCAACGCGCTCGCGCTCGCGCTCGCC 2297
 QY 360 AlaSerVal-----ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
 Db 2296 GTAGAGTCCACATCGTCCACACGTATACGCGACCTCCAGGTCCAGCTGATCGCCCC 2237
 QY 376 AsnGlyThrGlnTyValGlyAsnAspPhe---ThrSerProTyAsnAsnTrpAsp 394
 Db 2236 GACGCGACGGGTACACGCTCAAGTCTGACGCGACCGCGCGCAGTTCGGAC----- 2186
 QY 395 GlyArgAsnAsnValGluAsnValPhe---IleAsnAlaProGln-----SerGly 410
 Db 2185 -----AACATCAACACCGTACTCGGTGAACGCTCTCTCGGAGCGCGCCAACGCGC 2135
 QY 411 ThrTyThrIleGluVal 416
 Db 2134 ACGTGAACCTGCGGGTG 2117

RESULT 15

US-09-927-827-29
 ; Sequence 29, Application US/09927827
 ; Publication No. US20030036176A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bower, Stanley G.

Alignment Scores:			
Pred. No.:	6.96e-21	Length:	3743
Score:	283.00	Matches:	127
Percent Similarity:	37.53%	Conservative:	55
Best Local Similarity:	26.19%	Mismatches:	147
Query Match:	12.59%	Indels:	156
DB:	10	Gaps:	24
US-09-995-689A-1 (1-434) x US-09-927-827-29 (1-3743)			
Qy	21	GlyGlnGlyGlnIleValAlaValAlaAAspThrGlyLeu-----AspThrGly	36
Db	1502	GGCAGCGGCACGGTCGTGGCGGTGATTGATACCGCATCACCAGTCATCGGACCTCAAC	1561
Qy	37	ArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeu	56
Db	1562	GCCAAACATCTCTGGGGCTACGACTTCATCAGCGATCGACACCGCAGCGATGGCAAC	1621
Qy	57	GlyArgThrAsnAsnAlaAsnAsp-----	64
Db	1622	GCCCGTGACAGACACGCCCGCGACGAAGGCGACTGGTACCCGCCACAGNATCGCGGCGC	1681
Qy	65	-----ThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGly	78
Db	1682	GGCATTCCCGCGCGCAGCTCCAGTGTGCACGGCACCCCATGTGGCGCGCAGCGTCGGGCA	1741
Qy	79	AsnGlySerThrAsnLysGlyMetAlaProGlnAla-----	90
Db	1742	GTGACCAACAAACACCACCGCGGTAGCCGCGCACCGCCTACCGCGCCAAAGTCTGTACCGGTG	1801
Qy	91	-----AsnLeuValPheGlnSerIleMetAsp	99
Db	1802	CGCGTGCTCGGCAAGTCGGTGGTGCCTGCTCGGATATCCCGACGCCATCGTCTGGGCGC	1861
Qy	100	SerGlyGlyGlyLeuGlyLeuProSerAsnLeuGln-----ThrLeuPheSerGln	117
Db	1862	TCCGGCGGCACCGTCAGCGGCATCCCGCCAAATGCTAACCCCGCGCAGGTGATCAACATG	1921
Qy	118	AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla	137
Db	1922	TGCTGTGGCGCGCGCGGTAGTGTGTGCACACACCATGCAGAAC---GCCATCAACGGTGGC	1978
Qy	138	TyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeu	157
Db	1979	GTGTGCGCGCGCAC-----ACGTGGTG	2002
Qy	158	PheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys	177
Db	2003	GTCGCGCGCGCAACAGATCGCTCCAATGTGTCCGGT---TCGTCGCGCGCAACTGCGCG	2059
Qy	178	AsnAlaIleThrValGlyAlaThrGluAsn-----LeuArgProSerPheGlySerTyr	195
Db	2060	AACGTGATTGCGGTGGCGCGCCACACCTTCGGCGGGCGGAAGCGCAGCTATTCCACTTC	2119
Qy	196	AlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArg	215

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 23:20:18 ; Search time 2368 Seconds
(without alignments)
5473.052 Million cell updates/sec

Title: US-09-985-689a-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSGLY.....EYQAYNPVGPQTFLAIWN 434

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US0985689/runat_10032004_112903_19733/app_query.fasta_1.583
-DB=EST -QPMF=fastap -SUFFIX=xt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0985689@cgn_1_11335@runat_10032004_112903_19733 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hrc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hrc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_pro:*
26: em_gss_rod:*
27: em_gss_phg:*
28: em_gss_vrl:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	246	10.9	1605	13	BQ622771 CC Contig
2	241	10.7	640	12	BJ353336 BJ353336
3	225.5	10.0	4198	11	AK029048 Mus muscu
4	224.5	10.0	594	12	BJ393752 BJ393752
5	218.5	9.7	532	29	AL492464 T. brucei
6	207.5	9.2	2141	13	BQ142519 Contig6 M
7	200	8.9	508	28	AC652212 Sheared D
8	199.5	8.9	601	12	BJ387574 BJ387574
9	195.5	8.7	1002	29	CNS06D58
10	195	8.7	771	14	CA320325 UI-M-FW0-
11	191	8.5	633	12	BJ369190 BJ369190
12	190	8.5	718	12	BI750157 F902_1090
13	185.5	8.3	574	29	TA155H10P
14	184.5	8.2	4662	11	BC060627 Mus muscu
15	182.5	8.1	614	9	AJ273402 AJ273402
16	181	8.1	716	28	BZ893395 HL2_0177
17	180	8.0	641	12	BJ393925 BJ393925
18	180	8.0	665	13	BQ770462 UI-M-F10-
19	179.5	8.0	3091	11	BC011275 Mus muscu
20	177.5	7.9	650	9	AJ274038 AJ274038
21	177.5	7.9	681	14	CB690041 CBST-54-B
22	177	7.9	675	14	CF727824 UI-M-HB0-
23	176	7.8	530	29	CNS01090
24	176	7.8	576	14	CD295943 Strp0691.
25	176	7.8	2121	28	BZ424995 100023066
26	172.5	7.7	580	9	AJ273745 AJ273745
27	172.5	7.7	583	9	AJ273947 AJ273947
28	172.5	7.7	593	9	AJ273918 AJ273918
29	172.5	7.7	601	9	AJ273921 AJ273921
30	171.5	7.6	573	14	CA937626 88V42b10.
31	170.5	7.6	601	9	AJ273050 AJ273050
32	170.5	7.6	712	14	CD311344 Strp0691.
33	170.5	7.6	1029	29	CNS071DWM
34	170	7.6	895	13	BQ216158 AGENCOURT
35	170	7.6	1572	11	AY107161 Zea mays
36	169.5	7.5	545	13	BU575479 TGESTzyb8
37	169.5	7.5	604	9	AJ273185 AJ273185
38	168.5	7.5	601	12	BJ365857 BJ365857
39	168.5	7.5	609	9	AJ272712 AJ272712
40	168.5	7.5	937	12	BG246418 602360428
41	168	7.5	794	14	CF737198 UI-M-HD0-
42	168	7.5	1050	29	CNS0780L
43	167.5	7.5	564	28	AQ651427 Sheared D
44	167.5	7.5	610	9	AJ274218 AJ274218
45	167	7.4	449	12	BJ359939 BJ359939

ALIGNMENTS

RESULT 1
BQ622771
LOCUS CC Contig67
DEFINITION Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
CDNA, mRNA sequence.
ACCESSION BQ622771
VERSION BQ622771.1 GI:21649940
KEYWORDS EST
SOURCE Conidiobolus coronatus (Delacroixia coronata)
ORGANISM Conidiobolus coronatus
Eukaryota; Fungi; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
REFERENCE 1 (bases 1 to 1605)

AUTHORS Freimoser, F.M., Screen, S., Hu, G. and St. Leger, R.J.
TITLE EST analysis of genes expressed by the zygomycete pathogen *Conidiobolus coronatus* during optimized secretion of proteins
JOURNAL Unpublished (2002)
COMMENT Contact: Freimoser, F. M.
 Department of Entomology
 University of Maryland
 4112 Plant Sciences Building, College Park, MD 20742, USA
 Tel: 301 405 16 13
 Fax: 301 314 92 90
 Email: ff34@mail.umd.edu.

FEATURES
 source
 1..1605
 /organism="Conidiobolus coronatus"
 /mol_type="mRNA"
 /strain="ARSEF 512"
 /db_xref="taxon:34488"
 /clone_lib="Conidiobolus coronatus ARSEF 512"
 /note="Vector: Unizap; Conidiobolus coronatus was grown in minimal medium supplemented with Manduca sexta cuticle and peptone for 18 hours. A cDNA library was constructed in the unidirectional Lambda vector Unizap."

ORIGIN

Alignment Scores:
 Pred. No.: 1,528-13 Length: 1605
 Score: 246.00 Matches: 85
 Percent Similarity: 43.9% Conservative: 43
 Best Local Similarity: 29.21% Mismatches: 83
 Query Match: 10.95% Indels: 80
 DB: 13 Gaps: 14

US-09-985-689A-1 (1-434) x B0622771 (1-1605)

QY 6 GlyIleValIysAlaAspValAlaGlnSerSer-----TyrGlyLeu----- 19
 DB 718 GGTGTTGCTAAGGCTATTGCCGTCACAGCAATGCTCCTCGGGCTCTGCTGCTGGT 777
 QY 20 -----TyrGlyGlnGly 23
 DB 778 CAACTGCTAAGCTCGGATCTGCTCTTACTCTTACACCATGCTGCTGCTCAAGGT 837
 QY 24 GlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHis 43
 DB 838 GTCACGTGTTTGGTTTATAGACTGCTGCTCAATGTCAGCCCAATGAC----- 885
 QY 44 GluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAla--- 62
 DB 886 -----TTCGGTGGTCGGTCCACT-----TGGGTACTAACACTGCTGGT 924
 QY 63 -----AsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
 DB 925 GGTAGCAACACTGATGCTCACGGTCACGGTACTCCTGCTGCTGCTATTGCTGGT-- 981
 QY 80 GlySerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet--- 98
 DB 982 -----ACCACCTATGGTGTGCCAAGAGCTAACATGTTGCCGTTAAGTCTTAGGT 1035
 QY 99 ---AspSerGlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
 DB 1036 GATGATGGCTCGGATCATACTCTCGAATTATCTCCGGTATTGACTGCGGTGTTAAG-- 1092
 QY 118 AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla 137
 DB 1093 ---CACTCTGCTGCCAAGAAGATTATCTCTATGAGTTTAGAGGTGGTAAAGACGATGCT 1149
 QY 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeu 157
 DB 1150 CTTACACT-----GCTGTTAACACGCTGTACCAAGGGAGTGTGCTGTTGTC 1200
 QY 158 PheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys 177
 DB 1201 ---GCTGCTGGTACGACAGAGATGCTTGTGGATACTCT---CCCGCTTCTGCTCCT 1254

QY 178 AsnAlaIleThrValGlyAlaThrGlu-----AsnLeuArgProSerPheGlySerTyr 195
 DB 1255 TCCGCCATTACCGTGGTGGCCACTGATGTCATGATAAAAGGCTTCATTCTTACTTC 1314
 QY 196 AlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArg 215
 DB 1315 GGTAGCTGTGTC----- 1326
 QY 216 IleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAla 235
 DB 1327 -----GATATCTTAGCTCTGCTGTCACATCTCTCCACC----- 1362
 QY 236 ProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSer 255
 DB 1363 -----TGAAGGGATCTAACACTGCCACCAACCATCTCTGCTGCTCTCT 1407
 QY 256 MetAlaThrProIleValAlaGlyAsnValAla 266
 DB 1408 ATGGCTTGCCCTCACATTGCTGTTAGCTGCT 1440

RESULT 2

LOCUS BJ395336 640 bp mRNA linear EST 08-MAR-2002
 DEFINITION BJ395336 Dictyostelium discoideum cDNA library, SP Dictyostelium
 discoideum cDNA clone dds38b16 5', mRNA sequence.
 ACCESSION BJ395336
 VERSION BJ395336.1 GI:19306422
 KEYWORDS EST.
 SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 REFERENCE 1 (bases 1 to 640)
 AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
 TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.

FEATURES

source
 1..640
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="dds38b16"
 /sex="mat A"
 /dev_stage="slug stage"
 /clone_lib="Dictyostelium discoideum cDNA library, SP"

ORIGIN

Alignment Scores:
 Pred. No.: 1,036-13 Length: 640
 Score: 241.00 Matches: 69
 Percent Similarity: 49.76% Conservative: 36
 Best Local Similarity: 32.70% Mismatches: 72
 Query Match: 10.73% Indels: 34
 DB: 12 Gaps: 9

US-09-985-689A-1 (1-434) x BJ395336 (1-640)

QY 19 LeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg--- 37
 DB 22 TTAAGAGGTAAGGTCAGATATTGATGTTGCTGATCTGGTTTAGATGGTAGCCATTGT 81
 QY 38 -----AsnAspSer-----SerMetHisGluAlaPheArgGly 48
 DB 82 TTCTTTTCATGATCAAGATCATCAATCAATTTAATCAAGTGAATCAAAATCATAGAAA 141

SVTKAAWEGIAQOIHIMITVASPAETELHSGABHTSTVKLPIKIKIIPPPSRKVL
 WQYHNLRYPPGYPRDLNRMKNDPLDNGHDVHTNFRDMYQHLRSMGYEVVLGAPF
 TCFDQVGTLLVDSREVEYPPPIAKLRDNDVGLSLVIFSDWNTYSVVRKYKVEDE
 NTRQWMDPTGGMALPALNELLSVNMKGRSDGLVEGVELANDHMYVASGCSIAKPEE
 DGVITQTFKQGLVLEKQETAVENVPIILGYOIPSEGGRIYVLDGSDNCLDSDHQ
 KDCFLWLDALLOYSYGVTPSPSLHSGNRQRPSPGAGLAPFERMEGHLHRYSKVLEA
 RLSPGSESGAWDIFPGIMPRYNOEVGQTFVFAFLGAMVALLAFVVFVQISAKSRPKR
 RRPRAKRQOLAQAHPARTPSV"

polyA_signal

4180..4185

/note="putative"

polyA_site

4198

/note="putative"

ORIGIN

Alignment Scores:

Pred. No.: 8, 97e-11 Length: 4198
 Score: 225.50 Matches: 116
 Percent Similarity: 41.32% Conservative: 65
 Best Local Similarity: 28.48% Mismatches: 160
 Query Match: 10.04% Indels: 97
 DB: 11 Gaps: 25

US-09-985-689A-1 (1-434) x AK029048 (1-4198)

QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnLeuValAla 27
 DB 1005 CTGCGAGGAGATGCTGCTGG---CAGATGGATACACAGGTGCTAATGTCCAGAGTTGCT 1061
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
 DB 1062 GTTTTGTACTGGGCTC-----AGTGAGAGCATCCGCAATTTTAAG 1103
 QY 48 GlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn----- 63
 DB 1104 AAT-----GTGAGGAGAGAACCAACTGGACCAATGACCGGACCGCTG 1145
 QY 64 ---AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThr 82
 DB 1146 GATGATGGCTAGGCATCGCACATTCGTTCAGCT---GTGATTGCCAGCATGAGGAG 1202
 QY 83 AsnLysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerGly 101
 DB 1203 TGCCAAAGGATTTGCTCCAGATGAGAGCTGCACATCTTCAGGGTCTTTTACCACAAAT--- 1259
 QY 102 GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
 DB 1260 -----CAGGTGCTTACACATCTGTTCTCGATGCTTCACTAT 1301
 QY 122 GlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyValTyrThrAsp 141
 DB 1302 GCATCTCTAAAGAGATGAGCTTCTCAACCTTAGCATCGGTGGGCCGCACTTCATGGAT 1361
 QY 142 SerArgAsnValAspAspTyr-----ValArgLysAsnAspMetThrIleLeuPheAla 159
 DB 1362 CATCCGTTTGTGACAAGGTGGGAATTAACAGCTAACATGTAATGATGTTCTGCT 1421
 QY 160 AlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 179
 DB 1422 ATTGGCAATGATGAGCTCTCATGGCACTCTGAATAACCTCGTCGATCAGATGATGATG 1481
 QY 180 IleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIle 199
 DB 1482 ATTGGAGTGGGTGGGATTGAC-----TTTGAAGATAACATC 1517
 QY 200 AsnHisValAlaGlnPheSerSerArgGly-----ProThrLysAsp 213
 DB 1518 -----GCTCGCTTTTCTTCCAGGGGAATGACTACTCTGGGAATTTACAGAGGCTAT 1568
 QY 214 GlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer 233
 DB 1569 GTCGTGTGAGCCCTGACATTGCT-----ACCTATGTTGCTGGAGTGGGGGTTC 1619

QY 234 LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGly 253
 DB 1620 GGTGTGAAGGGGGCTGC-----CGTGCACTCTCAGGG 1652
 QY 254 ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal 273
 DB 1653 ACCAGTGTCCGTTCCCGAGTGTCTGGGGCCGCACCTTGTGTAGTAAAGCACAGTACAG 1712
 QY 274 LysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla 293
 DB 1713 AAGGGGAGCTGGTG-----AATCTGCCAGTGTGAAGCAAGCTTTGATACGCTCAGCC 1766
 QY 294 AlaAspIleGlyLeuGlyTyrProAsnGlyAsn-----GlnGlyTrpGlyArgVal 310
 DB 1767 CGGAGACTT-----CCTGGGTCAACATGTTCCGACAGGTGATGCCAAGTGTG 1814
 QY 311 ThrLeuAspLysSerLeuAsnVal---AlaTyrValAsnGluSerSerSerLeuSerThr 329
 DB 1815 GATCTGCTCGAGCTATCATGATCTCAGCAGCTATAAACCGCAGCAAGCTGAGTCTCT 1874
 QY 330 Ser-----GlnLysAlaThrTyrSerPheThrAlaThrAlaGlyLysProLeu 345
 DB 1875 AGCTACATCAGCTGAGTGTGCTCATATGTGG---CCCTACTGTCTCCAGCT--- 1928
 QY 346 LysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVal 365
 DB 1929 -----ATCTACTATGAGGAATGCCA-----ACAATCGTTAATGTCAACCTCCTC 1973
 QY 366 AsnAspLeuAspLeu-----ValIleThrAlaProAsnGlyThrGlnTyrValGly 382
 DB 1974 AATGGCATGGGCTCAGAGAAAGATTGGTAAAGCTGAGTGGCGACCCCTATTATTA--- 2030
 QY 383 AsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnVal 402
 DB 2031 -----CCACAGATGGA-----GACACATTCGAGTGGCC 2060
 QY 403 Phe-----IleAsnAlaProGlnSerGlyThrTyrThrIleGluVal 416
 DB 2061 TTCTCCTACTCCTCAGTGTGTGGCCCTGTCAGGTACCTTGCCTTCCCAATT 2114

RESULT 4

BJ393752

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BJ393752 594 bp mRNA linear EST 08-MAR-2002
 BJ393752 Dictyostelium discoideum cDNA library, SF Dictyostelium
 discoideum cDNA clone dds32b16 5', mRNA sequence.

BJ393752

BJ393752.1 GI:19304838

Dictyostelium discoideum

Dictyostelium discoideum

Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

1 (bases 1 to 594)

Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.

Full length cDNA of Dictyostelium discoideum at the slug stage

Unpublished (2002)

Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshinigenes.nig.ac.jp.

Location/Qualifiers

1..594

/organism="Dictyostelium discoideum"

/mol_type="mRNA"

/strain="AX4"

/db_xref="taxon:44689"

/clone="dds32b16"

/sex="mat A"

/dev_stage="slug stage"

/clone_lib="Dictyostelium discoideum cDNA library, SF"

FEATURES

source

ORIGIN

Alignment Scores: 4.2e-12 Length: 594
 Pred. No.: 224.50 Matches: 59
 Score: 224.50
 Percent Similarity: 50.80% Conservative: 36
 Best Local Similarity: 31.55% Mismatches: 59
 Query Match: 9.99% Indels: 33
 DB: 12 Gaps: 9

US-09-985-689A-1 (1-434) x BJ393752 (1-594)

QY 49 LysileThralaleuThralaleuGlyArgThrasnAsnAlaAspThrasnGlyHis 68
 Db 21 AAAGTTGTAACCTATATACCATCAACACGACGATAGTAAAGTGGTGCAC 80
 QY 69 GlyThrHisValalaglySerValleuGly-----AsnGlySer 81
 Db 81 GGTACACATATTGTGGTTCTGCACAGGTCTCCAGAGGATTCTTCAGTTAATATTCA 140
 QY 82 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 101
 Db 141 TCATTAGTGGTCTTGCACACTGATCAAGATTGCATTC-----TTTGATTGGCA 191
 QY 102 GlyGlyLeuGlyGlyLeu-----ProSerAsnLeuGlnThrLeuPheSerGlnAlaTyr 119
 Db 192 AGTGGTTTCATCAAGTTTCACACCTCCATCGGATTTGAAACCAATATATCAACCATATTAT 251
 QY 120 SerAlaGlyValaArgIleHisThrAsnSerTyrGlyVala-----AlaValAsn 135
 Db 252 GACGAGGTGCAGAGTGTGATTCCTGGGTTCTGTATCATAGTAGAGGGGTATACA 311
 QY 136 GlyAlaTyrThrAspSerArgAsnValAspAspTyrValArgLysAsn---AspMet 154
 Db 312 GGTAGTTATTCATCAGACACTGTTCAATTGATGATTTCCCTTTCACATCCAGATTTC 371
 QY 155 ThrIleLeuPheAlaGlyAsnGluGlyProAsnGlyGlyThrIleSer-----Ala 172
 Db 372 ATCATCTTAGAGTGTCTGTAAAC-----AAGCAGCAATACCTATCATCTACTACT 422
 QY 173 ProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg-----189
 Db 423 CAATCCACTGCAAGAAATGTTATTACCGTTGGTGTCTCATCAACAATTCATGAAATAT 482
 QY 190 -----ProSerPheGlySerTyrAlaAspAsnIleAsn-----His 201
 Db 483 TTAACGTGATGGTCCAAATATATTAATTAATCAATCATCTGCGATATAAATCAAGAGTTA 542
 QY 202 ValAlaGlnPheSerSerArg 208
 Db 543 ATATGTGATTTCGATAGCAGA 563

RESULT 5
 TA319G10P 532 bp DNA linear GSS 13-DEC-2000
 LOCUS T. brucei sheared genomic DNA clone 319g10, forward sequence,
 DEFINITION genomic survey sequence.
 ACCESSION AL492464
 VERSION AL492464.1 GI:11867408
 KEYWORDS GSS.
 SOURCE Trypanosoma brucei
 ORGANISM Trypanosoma brucei
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma
 1 (bases 1 to 532)
 REFERENCE Hall, N., Bowman, S., Lenhard, N.J., Doggett, J., Atkin, R.,
 AUTHORS Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
 Melville, S.E., Rajandream, M.A. and Barrell, B.G.
 TITLE Direct Submission
 JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
 project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
 Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
 nhls@sanger.ac.uk
 COMMENT Constructed at the Institute for Genomic Research (TIGR),

Rockville, MD. Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
 to give a tight size distribution (4 kb). The v + i method used for the library construction is
 described in detail in Smith, H. and Venter, J.C. (Making small
 insert libraries for whole genome shotgun sequencing projects. In
 Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999).

Email: nhls@sanger.ac.uk

Details of T. brucei sequencing at the Sanger Centre are available
 at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES

source
 1. 532
 /organism="Trypanosoma brucei"
 /mol_type="genomic DNA"
 /strain="TREU927"
 /db_xref="taxon:5691"
 /clone="319g10"

ORIGIN

Alignment Scores: 1.4e-11 Length: 532
 Pred. No.: 218.50 Matches: 61
 Score: 218.50
 Percent Similarity: 48.37% Conservative: 28
 Best Local Similarity: 33.15% Mismatches: 68
 Query Match: 9.72% Indels: 27
 DB: 29 Gaps: 6

US-09-985-689A-1 (1-434) x TA319G10P (1-532)

QY 93 ValPheGlnSerIleMetAspSerGlyGlyLeuGly-----105
 Db 2 GTAATGCCAAGTATAATGTGTCCAGGGGGCGAAGATTCTTCCAGGGGTTGGGCTGCC 61
 QY 106 -----GlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyr 119
 Db 62 CATCCAGTCAGAGTGTCTCCCTCCACGAGTTACTCAATTATTCGCCGGTATAT 121
 QY 120 SerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyr 139
 Db 122 GGGCGTGGAGCCGCTGTGTCTCAAACTCGTGGGTTTGTGCTCCCTCCGAGTATCT 181
 QY 140 ThrAspSerArgAsnValAspAspTyrValArgLys---AsnAspMetThrIleLeuPhe 158
 Db 182 GCTGTGAAAGGATATGATGATGTTCGAGTAGTATGACAGTGGTACTATATCTTC 241
 QY 159 AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsn 178
 Db 242 TCCACTGGCAACAGTATCCAGATGCG-----CTAATGACTCCGTCGTGTAAGAAC 295
 QY 179 AlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsn 198
 Db 296 GTGATGTGCGTGGGTTCACACAAAACGTCG-----TTTGACGCTTCGAAAGAC--- 343
 QY 199 IleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysPro 218
 Db 344 -----ATTGTTCTTCGTTTCTTCGATGTGCCACATACGACGATAGGATGAACCC 397
 QY 219 AspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer 238
 Db 398 GATCTTGTGCTCCCGGGAAGAGTGGTGTCTGTCTCTTTCTCTCTGCAAGCATCAGCT 457
 QY 239 SerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThr 258
 Db 458 AAA-----CAATGTAAGTGTGGCGCAAGCGGNTTCATCATGATGGCAACT 502
 QY 259 ProIleValAla 262
 Db 503 GCGGCCGTCGCG 514

RESULT 6
 BQ1425.9/c
 LOCUS BQ142519 2141 bp mRNA linear EST 24-APR-2002

Saccharomycetales; Saccharomycetaceae; Zygosaccharomyces.

REFERENCE
AUTHORS
1 (bases 1 to 1002)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekaiia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)

TITLE
JOURNAL
MEDLINE
PUBMED
20584718
11152876
11152883

REFERENCE
AUTHORS
2 (bases 1 to 1002)
de Montigny,J., Straub,M., Potier,S., Tekaiia,F., Dujon,B.,
Wincker,P., Artiguenave,F. and Souciet,J.
Genomic exploration of the hemiascomycetous yeasts: 8.
Zygosaccharomyces rouxii
FEBS Lett. 487 (1), 52-55 (2000)

TITLE
JOURNAL
MEDLINE
PUBMED
20584718
11152883

REFERENCE
AUTHORS
3 (bases 1 to 1002)
Genoscope
Direct Submission
Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Creneau, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
source
1..1002
/organism="Zygosaccharomyces rouxii"
/mol_type="genomic DNA"
/strain="CBS 732"
/db_xref="taxon:4956"
/clone="AR0A010H04"
/clone_lib="AR0A"
/note="end : 13"
<2..>979
/note="similar to Saccharomyces cerevisiae ORF YCR045c [
similarity to serin proteases]"
/evidence=not_experimental

misc_feature
2..>979
/note="similar to Saccharomyces cerevisiae ORF YCR045c [
similarity to serin proteases]"
/evidence=not_experimental

ORIGIN
Alignment Scores:
Pred. No.: 8,498-09 Length: 1002
Score: 195.50 Matches: 81
Percent Similarity: 41.90% Conservative: 38
Best Local Similarity: 28.52% Mismatches: 102
Query Match: 8,70% Indels: 63
DB: 29 Gaps: 14

US-09-985-689a-1 (1-434) x CNS06D6B (1-1002)

QY 14 GlnSerSerTyrGlyLeuTyr-----GlyGlnGlyGlnLeuValala 27
Db 278 GAAGAGAGTTTCAATTATTACTATTATAATGGCACAGGGTAGAATATCAATGCATAT 337

QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 338 ATWATTGATCTAGTGAATC-----TATAAGAACATPAAGGATCTTCTGT 379

QY 48 GlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAla---AsnAspThrAsn 66
Db 380 GGACGT-----GCCATATTCGGCCGAGACTTTACAGGGGAGGTCCTGGTGTGTCGTAAT 433

QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLysGlyMet 86
Db 434 GGACATGGAACTCAGCTGCTGCATTGG-----GGATCGAGCAACTTTGCTGT 484

QY 87 AlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGlyGlyGlyLeuGlyGly 106
Db 485 GCTAAAAAGGTGAATTAAATTCAGGTAAAGGTTTGAACATAGAGGC-----532

QY 107 LeuProSerAsnLeuGlnThrLeuPheSer-----GlnAlaTyr 119
Db 533 ---CAGGCAATTTAACACAGTAATTAGCGCTGTGAATTTGCCGTGATCATTTGATAA 589

QY 120 SerAlaGlyAlaArg-----IleHisThrAsnSerTyrGlyValAlaValAsnGlyAla 137
Db 590 TCATCGGGCAAAAAGGGTTCGCTGCCAATCTTTTCATTTGGGTGGCGTCAGAACAGTGT 649

QY 138 TyrThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeu 157
Db 650 ATAAATCAAGCAATTAATA-----GCTGCTCATGAAGCAGCGCTCATTTATTGTA 697

QY 158 PheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys 177
Db 698 GTTGCAGCAGGTAATTTCTAATTAAT---GCATGKGAATAGCCCTCTCTGCACCT 754

QY 178 AsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAsp 197
Db 755 GAGGCTATTACTGTAGCGCC-----TTTGATGAT 784

QY 198 AsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLys 217
Db 785 CGTACAGATCACTATTCAAAAATTTAGTAATTTGGGTCCCATGT-----826

QY 218 ProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAsp 237
Db 827 GTTGATATTTTGCATCAGCGTAAAGTTAAATCATTTG---TCTGCATTCACCT---880

QY 238 SerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAla 257
Db 881 -----CATAAACCAATTCATTTTCAGGTACTTCGATGGCT 916

QY 258 ThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGly 277
Db 917 TCACCAAGTGTTACTGAGCTTGTGCTATATTATTGGATAAGCGGTTGAACACAGAAAT 976

QY 278 IleThrProLys 281
Db 977 ATTAAGGCTAAG 988

RESULT 10
CA320325
LOCUS
DEFINITION
UI-M-FW0-cby-d-23-0-UI.r1 NIH_BMAP_FW0 Mus musculus cDNA clone
IMAGE:6816072 5', mRNA sequence.
CA320325
VERSION
CA320325.1 GI:24538449
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 771)
NIH-MGC http://mgi.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.

Location/Qualifiers

FEATURES

source

1. .771

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:6816072"

/tissue_type="whole brain"

/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"

/lab_host="PH10B (T1 phage resistant)"

/clone_lib="NIH BMAP FWO"

/note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

Alignment Scores:

Pred. No.: 6,14e-09 Length: 771

Score: 195.00 Matches: 82

Percent Similarity: 41.81% Conservative: 38

Best Local Similarity: 28.57% Mismatches: 109

Query Match: 8.68% Indels: 58

DB: 14 Gaps: 13

US-09-985-689A-1 (1-434) x CA320325 (1-771)

QY 8 ValLysAlaAspValAlaGlnSerSerTyGlyLeuTyGlyGlnGlyValAla 27
Db 40 CTGACGAGAGTGTGTGG---CAGATGGGATACACAGGTGCTAATCTCAGATTGCT 96
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 97 GTTTTGTACTGGGCTC-----AGTGAGAAGCATCCGCATTATTAAG 138
QY 48 GlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsn----- 63
Db 139 AAT-----GTGAGGAGAGAGACCACTGGACCAATGAGCGGACCCTG 180
QY 64 ---AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThr 82
Db 181 GATGATGGGTAGCCATGCCATTCGTTGAGGT---GTGATTGCCAGCATGAGGGAG 237
QY 83 AsnLysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerGly 101
Db 238 TGCCAGGATTGCTCCAGATGCGAGCTGCACATCTTCAGGTCTTTACCAACAAAT--- 294
QY 102 GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTySerAla 121
Db 295 -----CAGGTGTCTTACACATCTTGGTTCTGGATGCTTCAACTAT 336
QY 122 GlyAlaArgIleHisThrAsnSerTPGlyAlaAlaValAsnGlyAlaTyThrThrAsp 141
Db 337 GCATCTCTAAGAAGATGGACGTTCTCAACCTTAGCATCGGTGGGCCGACATTCATGGAT 396
QY 142 SerArgAsnValAspAspTyTyr-----ValArgLysAsnAspMetThrIleLeuPheAla 159
Db 397 CATCCGTTTGTGACAGGTTGGGAATTAAACAGCTAAACAATGTAATTATGTTTCTGCT 456

QY 160 AlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 179
Db 457 ATTGGCAATGATGACCTCTCTATGGCACTCTGAATACCTGCTGATCAGATGGATGTG 516
QY 180 IleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIle 199
Db 517 ATTGGAGTGGGTGGCATTGAC-----TTTGAAGATAACATC 552
QY 200 AsnHisValAlaGlnPheSerSerArgGly-----ProThrLysAsp 213
Db 553 -----GCTCGCTTTTCTCCAGGGGAATGACTACCTGGGAATTACACAGAGGCTAT 603
QY 214 GlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer 233
Db 604 GGTGCTGTGAAGCTGACATTGTC-----ACCTATGCTGCTGGAGTCCGGGTTC 654
QY 234 LeuAlaProAspSerSerPheTPAlaAsnHisAspSerLysTyAlaTyMetGlyGly 253
Db 655 GGTGTGAAAGGGGGCTGC-----CGTGCACTCTCAGGG 687
QY 254 ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal 273
Db 688 ACCAGTCTGCTCCCGCAGTGTGCTGGGGCGGTACCTTTAGTANGCAGACATGACAG 747
QY 274 LysAsnArgGlyIleThrPro 280
Db 748 AAGCGGAGCTGTGTAATCCT 768

RESULT 11

BJ369190

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2002)

Contact: Tadasu Shin-i

Center for Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

Location/Qualifiers

source

1. .633

/organism="Dictyostelium discoideum"

/mol_type="mRNA"

/strain="AX4"

/db_xref="taxon:44689"

/clone="ddc49116"

/sex="mat A"

/dev_stage="Culmination stage"

/clone_lib="Dictyostelium discoideum cDNA library, CF"

ORIGIN

Alignment Scores:

Pred. No.: 1,11e-08 Length: 633

Score: 191.00 Matches: 57

Percent Similarity: 41.55% Conservative: 29

Best Local Similarity: 27.54% Mismatches: 49

Query Match: 8.50% Indels: 72

DB: 12 Gaps: 6

US-09-985-689A-1 (1-434) x BJ369190 (1-633)


```

QY 259 ProIleValAlaGlyAsnVal 265
Db 552 CYCCACATTGCGGACTTGTGC 572

RESULT 13
LOCUS TA315H10P 574 bp DNA linear GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 315h10, forward sequence,
ACCESSION AL490202
VERSION AL490202.1 GI:11866292
KEYWORDS Trypanosoma brucei
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei
REFERENCE 1 (bases 1 to 574)
AUTHORS Halli,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
Melville,S.E., Rajandream,M.A. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
Project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
rh@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTst 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/projects/T_brucei/.
LOCATION/Qualifiers
source
1. .574
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="315h10"

ORIGIN
Alignment Scores:
Pred. No.: 3 39e-08 Length: 574
Score: 185 50 Matches: 50
Percent Similarity: 48.20% Conservative: 30
Best Local Similarity: 28.25% Mismatches: 54
Query Match: 8.26% Indels: 43
DB: 29 Gaps: 7

US-09-985-689A-1 (1-434) x TA315H10P (1-574)

QY 18 GlyLeuTyGlyClnGlyClnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 63 GGATTGACGGTGTGCTGAGATAATAGGGTAGCGGATACGGGTATCGACTTT----- 116
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLyIleThrAlaLeuTyAlaLeuGly 57
Db 117 AACAGCTGTTTCTTCACGATCCA-----ATCAGAGGTGCGCTTTACCGGAGGTT 170
QY 58 ArgThrAsnAsn----- 61
Db 171 AACTATAACACACCGCAAAATCGTGTCAATTGGCCCGGTGTGACTTTCATCCGGGGGATTAC 230
QY 62 -----AlaAsnAspThrAnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 231 TTTCGTGGGGATGAGGAATAGTCTATGACGACGACGTCGTCAGTACCGACCGGGGAGT 290
QY 80 GlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 93

```

```

Db 291 GTTATTAGTAACGCGGTAAATGCCAAGTATATGGTGTCCCAAGGGCGGAAGATTTC 350
QY 94 PheGlnSerIleMetAspSerGlyGlyGlyLeuGly-----Gly 106
Db 351 TTCAGG-----GGTTGGGTGCCCATCCCGCAGCAGCTTGTCT 399
QY 107 LeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTySerAlaGlyAlaArgIleHis 126
Db 390 CTCCTCCACGACGCTTACTCAATATTATTCGCCGATATGGCGCTGGAGCCCTGTGTT 449
QY 127 ThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyThrThrAspSerArgAsnValAsp 146
Db 450 TCACACTCGTGGGGTTTGTGCTCCCTCCGACTATTCCTCTGCGAAGGATATGAT 509
QY 147 AspTyrValArgLys---AsnAspMetThrIleLeuPheAlaAlaGlyAsn 162
Db 510 GAGTTTCGAGTAGTATTGACGATGGCGCTACTTATCTCTCCACTGCGCAAC 560

RESULT 14
LOCUS BC060627 4662 bp mRNA linear HTC 19-NOV-2003
DEFINITION Mus musculus tripeptidyl peptidase II, mRNA (cDNA clone
IMAGS:6844459), containing frame-shift errors.
ACCESSION BC060627.1 GI:38148645
VERSION BC060627
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 4662)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,B.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.P., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.P., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Casavant,T.L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Usdin,T.B., Toshiyuki,S.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Casavant,T.L.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahy,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallos,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL MEDLINE
PUBMED 12477932
REFERENCE 2 (bases 1 to 4662)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgaaps-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
Thomas L. Casavant
Web site: http://genome.uiowa.edu
Contact: bento-soares@uiowa.edu, tom-casavant@uiowa.edu
Bonaldo,M.F., Akabogu,I., Bair,T., Bair,J., Crouch,K., Davis,A.,

```

Fishler, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K.,
Schaefer, T., Smith, C., Snir, E., Tack, D., Trout, K., Walters, J.,
Casavant, T., Soares, M.B.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: Plate: Row: Column: 0
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6678418
This clone has the following problem: frame shifted.

FEATURES

source

1..4662
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6844459"
/tissue_type="Brain, mouse, 13.5,14.5,16.5,17.5 dpc"
/clone_lib="NTH SWAP_FYO"
/lab_host="DH10B"
/note="Vector: pYX-ASC"

ORIGIN

Alignment Scores:
Pred. No.: 1.45e-06 Length: 4662
Score: 184.50 Matches: 101
Percent Similarity: 35.56% Conservative: 69
Best Local Similarity: 21.13% Mismatches: 155
Query Match: 8.21% Indels: 153
Gaps: 21

US-09-985-689A-1 (1-434) x BC060627 (1-4662)

QY 10 AlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAlaValAla 29
DB 803 GCTGAGATGCTCAATTACTCTGTGAACATTATGACGATGGCAACCTGCTCCATTGTG 862
QY 30 AspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLys 49
DB 863 ACCACCGGA----- 871
QY 50 IleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHisGly 69
DB 872 -----GGAGCTCATGGA 883
QY 70 ThrHisValAlaGlySerValLeuGlyAsn-----GlySerThrAsnLysGlyMet 86
DB 884 ACCCATGTAGCAAGTATACCCGACGGCATTTTCCAGAGAGCCCTGAAACCGAATGGAGTT 943
QY 87 AlaProGlnAlaAsnLeuValPhe-----GlnSerIleMetAsp 99
DB 944 GCTCCTGCTGCTCAATTCTATCCATTAAAGATTGGTGATACACGGCTAAGCACTATGGAA 1003
QY 100 SerGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAla--- 118
DB 1004 ACAGGCACAGGCCTC-----ATCAGAGCTATGATAGAAGTTATAAATCATCAATGATGAT 1057
QY 119 -----TyrSerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsn 135
DB 1058 CTTGTCACTACAGTTATGGAGCAACCACTATTGGCCCAATTTCTGGAGAAATT----- 1111
QY 136 GlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThr 155
DB 1112 -----TGTAAGTAATTAATGAAGCAGTATGGAACATACATAATACAAATT 1153
QY 156 IleLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaPro---Gly 174
DB 1154 TATGTTTCAAGCTGGGAATAATGTCATCCGCTTTCTACAGGGGTGTCAGAGGGA 1213
QY 175 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 194
DB 1214 ACTACATCCAGTGTATAGTGTGGAGCT----- 1243

QY 195 TyrAlaAspAsnIleAsnHisValAlaGlnPhe----- 205
DB 1244 TATGTTTCCCTGATATGATGTTGCAGAGATTTCATCTGAGAGAAACTCCCTGGCAAT 1303
QY 206 -----SerSerArgGlyProThrLysAspGlyArgIleLysProAspValMet 221
DB 1304 CAATATACATGCTCTCTAGAGGCCAAGTCTGATGAGCCCTCGGTGTGAGCATAGT 1363
QY 222 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr 241
DB 1364 GCACCAAGAGGTGCTATTGCTTCTGCTGCTAAC-----TGG 1399
QY 242 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 261
DB 1400 ACATTGAGGGGACTCAG-----CTAATGAATGGACATCAATGCTCTTCCCAATGCC 1453
QY 262 AlaGlyAsnValAla-----GlnLeuArgGluHisPheValLysAsnArgGly 277
DB 1454 TGTGTGGCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1513
QY 278 IleThrProLysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGly 297
DB 1514 CACTCAGTCAGAGAGACTCTAGAAAATACTCAATAAAGCTGACAATATAGAGTATT 1573
QY 298 LeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsn 317
DB 1574 GCC-----CAAGGACATGGAATATTTCAGGTGACAAAGCTTTATGAC 1615
QY 318 ValAlaTyrValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPhe 337
DB 1616 TACCTCAATCAAAAT-----ACATCATTTGCTAACAGATTAGTTT 1657
QY 338 ThrAlaThrAlaGlyLys-----ProLeuValIleSerLeu 349
DB 1658 ACAGTTACTGTTGGAAATAACCGTGTATCTACTCCGAGATCCCTGTCAGGTGGCTCT 1717
QY 350 ValTyrSerAsp-----AlaProAlaSerThrThrAlaSer 361
DB 1718 CTT---TCAGATCAGTGTGTTGGCATTGCGCTGATTTCAGAGAACACAGAACTCT 1774
QY 362 ValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrVal 381
DB 1775 GAAAAAATATCTTTTCCAGCTTCATTAGCTTTAACTTCA----- 1813
QY 382 GlyAsnAspPheThrSerProTyrAsnAspAsnTyr-----Asp 394
DB 1814 -----AATTCATCTGGTTCAGTGTCCAGCCCATTTGGAA 1849
QY 395 GlyArgAsnAsnValGluAsnValPheIleAsnAlaPro-----GlnSerGly 410
DB 1850 CTCATGAATCAGTGTCCGACATAAACAATACGTGTGGACCCCGAGGGGCTTAAGAGAGGG 1909
QY 411 ThrTyrThrIleGluValGlnAlaTyrAsnValPro-----ValGlyPro 425
DB 1910 TTACATTATACAGAGTATGTGGCTATGATATAGCATCCCAATGCAAGTCT 1963

RESULT 15

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

UNPUBLISHED (1999)

EST

EST

EST

EST

EST

EST

EST

EST

EST

EST

EST

EST

COMMENT Contact: Screen SE

Entomology
University of Maryland
4112 Plant Sciences Building, College Park, MD 20742, USA.
Location/Qualifiers

FEATURES

source
1..614
/organism="Metarhizium anisopliae"
/mol_type="mRNA"
/strain="ARSEP 2575"
/db_xref="taxon:5530"
/clone="Ma#948"
/clone_lib="Metarhizium anisopliae ARSEP 2575"
/note="Vector: Unizap; Metarhizium anisopliae was grown on insect cuticle for 24 hours. A cDNA library was constructed in the unidirectional Lambda vector, Unizap"

ORIGIN

Alignment Scores:

Pred. No.:	7.61e-08	Length:	614
Scores:	182.50	Matches:	67
Percent Similarity:	42.86%	Conservative:	38
Best Local Similarity:	27.24%	Mismatches:	88
Query Match:	8.12%	Indels:	53
DB:	9	Gaps:	12

US-09-985-689A-1 (1-434) x AJ273402 (1-614)

```
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 11 GGTGAGGTACTTCGGTATATATCATTCGACATCGTATTGAG-----52
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 53 GCCTCCACCCCGAGTTTGAGGTCGCGCCACITTTCTTAAGAGCTTCATCAGCGGTCAA 112
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 113 AAC---ACTGATGCCACGCGCATGGGACTCACTGCGCTGGTACCAATT-----GGT 160
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 161 AGCAAGACCTACGGTGTGGCCAAAAGCTAAGCTCTATGGTGTCAAGGTTCCTTGACAAC 220
Qy 101 -----GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAla 118
Db 221 CAGGCGCAGTGTCTCTACTCCGTATCATCATGTCATGAGTACGTGTCACAGGACTCC 280
Qy 119 TyrSerAlaGly-----AlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsn 135
Db 281 AAGACCGCGGCTCCCAACGCGGCCATTGCTTCCATGAGCCTGGGA-----328
Qy 136 GlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThr 155
Db 329 GGTGGTACTCGGGTCGCTCAACCAAGGTGCTGCTGTTGTCATTCCTGGTGCTTC 388
Qy 156 IleLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThr 175
Db 389 CTTCCGCTGCGCGCTGCGACGATACCGGATGCCCAACACCTCT---CCCGCTTCC 445
Qy 176 AlalysAsnAlaIleThrValGlyValAlaThr-----GluAsnLeuArgProSerPheGly 193
Db 446 GAGCCTTCGCTGCATCTGTTGGTCTCTGCGGAAATGACAGCGCATCTTCCTCTTC 505
Qy 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAsp 213
Db 506 AACTAC-----511
Qy 214 GlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer 233
Db 512 GGCAGAGTT---GTCGATATTTTCCTCTGTTAGCAATGTTCTTTCACC-----559
Qy 234 LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGly 253
Db 253
```

Db 560 -----TGGATTGGT-----GGCCGCACAAACACCATCTCTGGT 592

Qy 254 ThrSerMetAlaThrPro 259

Db 593 ACCTNCAATGGCTACTCCC 610

Search completed: March 16, 2004, 01:15:53
Job time : 2379 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:46:53 ; Search time 59 Seconds
(without alignments)
2078.400 Million cell updates/sec

Title: US-09-985-689a-1-COPY
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVOAYNPVGPQTFSLAIVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : A_Geneseq_29Jan04.*
- 1: Geneseqp1980s.*
 - 2: Geneseqp1990s.*
 - 3: Geneseqp2000s.*
 - 4: Geneseqp2001s.*
 - 5: Geneseqp2002s.*
 - 6: Geneseqp2003as.*
 - 7: Geneseqp2003bs.*
 - 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2242	99.8	434	5	AAW50080 Bacillus
2	2242	99.8	640	2	AAV17090 Bacillus
3	2237	99.5	640	2	AAV17091 Bacillus
4	2186	97.3	434	5	AAW50081 Bacillus
5	2178	96.9	639	2	AAV17089 Bacillus
6	2150	95.7	639	2	AAV17087 An alkali
7	2150	95.7	640	2	AAV17088 An alkali
8	2138	95.1	434	5	AAW50085 Bacillus
9	2120.5	94.4	433	5	AAW50086 Bacillus
10	2120.5	94.4	641	2	AAW89547 Bacillus
11	2071	92.2	434	5	AAW50090 Bacillus
12	1933.5	88.7	433	5	AAW50084 Bacillus
13	1939.5	88.5	433	5	AAW50082 Bacillus
14	1984.5	88.3	636	2	AAW89548 Bacillus
15	1982.5	88.2	433	5	AAW50083 Bacillus
16	1981.5	88.2	433	2	AAW26274 Alkali-pr
17	1981.5	88.2	433	2	AAW61495 Modified
18	1981.5	88.2	433	2	AAW95698 Bacillus
19	1981.5	88.2	433	3	AAV69207 Amino aci
20	1981.5	88.2	433	3	AAV44619 Bacillus
21	1581.5	70.4	345	2	AAW62230 Subtilase
22	1581.5	70.4	345	2	AAV21654 Subtilase
23	447.5	19.9	659	2	AAW24121 Thermococ
24	447.5	19.9	659	2	AAW94840 WO9856926
25	409	18.2	412	2	AAW94836 Hyperther

26	409	18.2	522	2	AAW24122 Pyrococcus
27	409	18.2	522	2	AAW94838 Hyperther
28	409	18.2	654	2	AAW24129 Pyrococcus
29	409	18.2	654	2	AAW94841 Hyperther
30	396	17.6	659	2	AAW24123 Protease.
31	372	16.6	545	4	ABO09483 T. yonsei
32	366	16.3	1079	6	ABO09483 T. yonsei
33	366	16.3	1079	6	ABO09483 T. yonsei
34	348	15.5	734	2	AAW13666 Fragment.
35	348	15.5	734	2	AAW13667 Streptomy
36	348	15.5	823	2	AAW13668 DhpA-mel
37	336	15.0	1237	6	ABU11343 Protein e
38	311.5	13.9	806	2	ABU27481 RP-III re
39	306.5	13.6	519	6	ABP76735 Streptomy
40	306.5	13.6	19938	6	ABP76678 Streptomy
41	305.5	13.6	903	2	AAW87007 Hyperther
42	305.5	13.6	1398	2	AAW87008 Protease.
43	305.5	13.6	1398	2	AAW24124 Pyrococcus
44	305.5	13.6	1398	2	AAW94839 WO9856926
45	296.5	13.2	699	2	AAV08471 F. balust

ALIGNMENTS

RESULT 1
AAM50080
ID AAM50080 standard; protein; 434 AA.
XX AC AAM50080;
XX DT 12-AUG-2002 (first entry)
XX DE Bacillus sp KSM-KP43 alkaline protease protein fragment.
XX KW Alkaline protease; detergent; laundry; bleaching; dishwasher.
XX OS Bacillus sp.
XX PN EP1209233-A2.
XX PD 29-MAY-2002.
XX PF 22-NOV-2001; 2001EP-00127851.
XX PR 22-NOV-2000; 2000JP-00355166.
XX PR 12-APR-2001; 2001JP-00114048.
XX PA (KAOS) KAO CORP.
XX PI Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
XX PI Okuda M, Saeki K;
XX DR WPI; 2002-437518/47.
XX PT New modified alkaline proteases useful in detergent compositions.
XX PS Claim 1; Page 10-11; 25pp; English.
XX CC This invention describes novel Bacillus sp. alkaline proteases useful in
CC detergent compositions, especially in laundry, bleaching or automatic
CC dishwasher detergents. The novel proteases have an increased detergency %
CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
CC sequence represents a fragment of the alkaline protease KP43 from
CC Bacillus sp strain KSM-KP43 which is used to create the modified protease
CC represented in AAM50090
XX SQ Sequence 434 AA;

Query Match 99.8%; Score 2242; DB 5; Length 434;
Best Local Similarity 99.8%; Pred. No. 7,2e-152;
Matches 433; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
 Db 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
 QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
 Db 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
 QY 121 AGARIHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 Db 121 AGARIHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTIFLSARSSLAPDSSF 240
 Db 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTIFLSARSSLAPDSSF 240
 QY 241 WANHDSKYAYMGTSWATPIVAGNVAQLREHFVKRGITPKPSLLKAALIGAADIGLGY 300
 Db 241 WANHDSKYAYMGTSWATPIVAGNVAQLREHFVKRGITPKPSLLKAALIGAADIGLGY 300
 QY 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTQKATYSFTATAGKPLKISLVMSDAPASTTA 360
 Db 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTQKATYSFTATAGKPLKISLVMSDAPASTTA 360
 QY 361 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDNDGNNVENVFINAPOSGTITIEVQAYN 420
 Db 361 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDNDGNNVENVFINAPOSGTITIEVQAYN 420
 QY 421 VPVGPOTFSLAIVN 434
 Db 421 VPVGPOTFSLAIVN 434

RESULT 2

AAV17090
 ID AAV17090 standard; protein; 640 AA.

XX AC AAV17090;
 XX DT 20-MAR-2003 (revised)
 XX DT 21-JUL-1999 (first entry)
 XX DE Bacillus alkaline protease.
 XX KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 XX KW washing composition; oxidising agent.
 XX OS Bacillus sp.

XX PN WO9918218-A1.
 XX PD 15-APR-1999.
 XX PF 07-OCT-1998; 98WO-JP004528.
 XX PR 07-OCT-1997; 97JP-00274570.
 XX PA (KAOS) KAO CORP.

XX PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 XX PI Shikata S, Nomura M;
 XX DR WPI; 1999-287736/27.
 XX DR N-PSDB; AAX37278.

XX PT Alkali protease from Bacillus used in washing powders.

XX PS Disclosure; Page 58-63; 71pp; Japanese.

XX CC The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it

CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease. (Updated
 CC on 20-MAR-2003 to correct DR field.)

XX Sequence 640 AA;

QY Query Match 99.8%; Score 2242; DB 2; Length 640;
 Db Best Local Similarity 99.8%; Pred. No. 1.2e-151;
 Mismatches 433; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
 Db 207 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 266
 QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
 Db 267 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 326
 QY 121 AGARIHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 Db 327 AGARIHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
 QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTIFLSARSSLAPDSSF 240
 Db 387 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTIFLSARSSLAPDSSF 446
 QY 241 WANHDSKYAYMGTSWATPIVAGNVAQLREHFVKRGITPKPSLLKAALIGAADIGLGY 300
 Db 447 WANHDSKYAYMGTSWATPIVAGNVAQLREHFVKRGITPKPSLLKAALIGAADIGLGY 506
 QY 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTQKATYSFTATAGKPLKISLVMSDAPASTTA 360
 Db 507 PNGNQGWGRVTLDKSLNVAYVNESSLSSTQKATYSFTATAGKPLKISLVMSDAPASTTA 566
 QY 361 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDNDGNNVENVFINAPOSGTITIEVQAYN 420
 Db 567 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDNDGNNVENVFINAPOSGTITIEVQAYN 626
 QY 421 VPVGPOTFSLAIVN 434
 Db 627 VPVGPOTFSLAIVN 640

RESULT 3

AAV17091
 ID AAV17091 standard; protein; 640 AA.

XX AC AAV17091;

XX DT 20-MAR-2003 (revised)
 XX DT 21-JUL-1999 (first entry)

XX DE Bacillus alkaline protease.

XX KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 XX KW washing composition; oxidising agent.

XX OS Bacillus sp.

XX PN WO9918218-A1.

XX PD 15-APR-1999.

XX PF 07-OCT-1998; 98WO-JP004528.

XX PR 07-OCT-1997; 97JP-00274570.

XX PA (KAOS) KAO CORP.
 XX DE Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX WPI; 1999-287736/27.
 DR N-PSDB; AAX37279.
 XX PT Alkali protease from Bacillus used in washing powders.
 XX PS Disclosure; Page 63-68; 7lpp; Japanese.
 XX CC The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidizing agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease. (Updated
 CC on 20-MAR-2003 to correct DR field.)
 XX SQ Sequence 640 AA;
 Query Match 99.6%; Score 2237; DB 2; Length 640;
 Best Local Similarity 99.5%; Pred. No. 2.7e-151;
 Matches 432; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 NDVARGIVKADVAQSSYGLYGQGI VAVADTGLDGRNDSMHEAFRGKITALYALGRN 60
 DB 207 NDVARGIVKADVAQSSYGLYGQGI VAVADTGLDGRNDSMHEAFRGKITALYALGRN 266
 QY 61 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLQTLFSQAYS 120
 DB 267 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLQTLFSQAYS 326
 QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 327 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
 QY 181 TVGATENLRPSFGSYADNINHVAQFSRGPTKDGRIKPDVMAPTTILSARSLAPDSSF 240
 DB 387 TVGATENLRPSFGSYADNINHVAQFSRGPTKDGRIKPDVMAPTTILSARSLAPDSSF 446
 QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 DB 447 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 506
 QY 301 PNGNQGWGRVTLDKSLNVAIVYNESSLSSTQKATYFTATAGKPLKISLWSDAPASTTA 360
 DB 507 PNGNQGWGRVTLDKSLNVAIVYNESSLSSTQKATYFTATAGKPLKISLWSDAPASTTA 566
 QY 361 SVTLVNDLNLVITAPNGTQVVGNDFTSPYNDNDGRNNVENVFINAPQSGTYTIEVQAYN 420
 DB 567 SVTLVNDLNLVITAPNGTQVVGNDFTSPYNDNDGRNNVENVFINAPQSGTYTIEVQAYN 626
 QY 421 VPVGQPTFSLAIVN 434
 DB 627 VPVGQPTFSLAIVN 640
 RESULT 4
 AAM50081
 ID AAM50081 standard; protein; 434 AA.
 XX
 AC
 XX AAM50081;
 XX

DT 12-AUG-2002 (first entry)
 XX DE Bacillus sp KSM-KP9860 alkaline protease protein fragment.
 KW Alkaline protease; detergent; laundry; bleaching; dishwasher.
 XX OS Bacillus sp.
 PN EPI209233-A2.
 XX 29-MAY-2002.
 PD
 XX 22-NOV-2001; 2001EP-00127851.
 XX 22-NOV-2000; 2000JP-00355166.
 PR 12-APR-2001; 2001JP-00114048.
 XX (KAOS) KAO CORP.
 PA Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
 PI Okuda M, Saeki K;
 XX WPI; 2002-437518/47.
 DR New modified alkaline proteases useful in detergent compositions.
 XX Claim 5; Page 12-13; 25pp; English.
 PS
 XX CC This invention describes novel Bacillus sp. alkaline proteases useful in
 CC detergent compositions, especially in laundry, bleaching or automatic
 CC dishwasher detergents. The novel proteases have an increased detergency
 CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
 CC sequence represents a fragment of the alkaline protease KP9860 from
 CC Bacillus sp strain KSM-KP9860 described in the method of the invention
 XX SQ Sequence 434 AA;
 Query Match 97.3%; Score 2186; DB 5; Length 434;
 Best Local Similarity 96.3%; Pred. No. 7.2e-148;
 Matches 418; Conservative 14; Mismatches 2; Indels 0; Gaps 0;
 QY 1 NDVARGIVKADVAQSSYGLYGQGI VAVADTGLDGRNDSMHEAFRGKITALYALGRN 60
 DB 1 NDVARGIVKADVAQSSYGLYGQGI VAVADTGLDGRNDSMHEAFRGKITALYALGRN 60
 QY 61 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLQTLFSQAYS 120
 DB 61 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLQTLFSQAYS 120
 QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 QY 181 TVGATENLRPSFGSYADNINHVAQFSRGPTKDGRIKPDVMAPTTILSARSLAPDSSF 240
 DB 181 TVGATENLRPSFGSYADNINHVAQFSRGPTKDGRIKPDVMAPTTILSARSLAPDSSF 240
 QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 DB 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 QY 301 PNGNQGWGRVTLDKSLNVAIVYNESSLSSTQKATYFTATAGKPLKISLWSDAPASTTA 360
 DB 301 PNGNQGWGRVTLDKSLNVAIVYNESSLSSTQKATYFTATAGKPLKISLWSDAPASTTA 360
 QY 361 SVTLVNDLNLVITAPNGTQVVGNDFTSPYNDNDGRNNVENVFINAPQSGTYTIEVQAYN 420
 DB 361 SVTLVNDLNLVITAPNGTQVVGNDFTSPYNDNDGRNNVENVFINAPQSGTYTIEVQAYN 420
 QY 421 VPVGQPTFSLAIVN 434
 DB 421 VPVGQPTFSLAIVN 434

RESULT 5
 AAY17089
 ID AAY17089 standard; protein; 639 AA.
 XX
 AC AAY17089;
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE Bacillus alkaline protease.
 XX
 KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent.
 XX
 OS Bacillus sp.
 XX
 PN WO9918218-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX
 PR 07-OCT-1997; 97JP-00274570.
 XX
 PA (KAOS) KAO CORP.
 XX
 PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX
 DR WPI; 1999-287736/27.
 DR N-PSDB; AAX37277.
 XX
 PT Alkali protease from Bacillus used in washing powders.
 XX
 PS Disclosure; Page 53-58; 71pp; Japanese.
 XX
 CC The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease. (Updated
 CC on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 639 AA;
 Query Match 96.9%; Score 2178; DB 2; Length 639;
 Best Local Similarity 96.1%; Pred. No. 4.4e-147;
 Matches 417; Conservative 14; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NDVARGIVKADYQAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTN 60
 DB 206 NDVARGIVKADYQAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTN 265
 QY 61 NANDTNGHGHVAGSVLGNSTNGKMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSAQYS 120
 DB 266 NANDTNGHGHVAGSVLGNSTNGKMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSAQYS 325
 QY 121 AGARHTNSGVAAGVNGAYTTDSRNVDDYVRKNDMTILPAAGNEGNGGTTISAPGAKNAI 180
 DB 326 AGARHTNSGVAAGVNGAYTTDSRNVDDYVRKNDMTILPAAGNERNGGTTISAPGAKNAI 385
 QY 181 TVGATENLRFSGSYADNINHVAFSSRGPTKDGRIKPDVMAAGFTILSARSSLPDSSF 240
 DB 386 TVGATENLRFSGSYADNINHVAFSSRGPTKDGRIKPDVMAAGFTILSARSSLPDSSF 445

241 WANHDSKYAYMGTSMTATPIVAGNVAQOLREHFVKNRGITPKPSLLKAAALTAGADIGLY 300
 DB 446 WANHDSKYAYMGTSMTATPIVAGNVAQOLREHFVKNRGITPKPSLLKAAALTAGADIGLY 505
 QY 301 PNGNQGWGRVTLDKSLNVAYVNESSLSLSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 360
 DB 506 PNGNQGWGRVTLDKSLNVAYVNESSLSLSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 565
 QY 361 SVTLVNDLNLVITAPNGTOYVGNDFSPYNDNDGNNVENVFNAPQSCTYTIIEVOAYN 420
 DB 566 SVTLVNDLNLVITAPNGTRYVGNDFSAFFNDNNDGNNVENVFNAPQSCTYTIIEVOAYN 625
 QY 421 VPVGPQTFSLAIVN 434
 DB 626 VPVGPQNFSLAIVN 639

RESULT 6
 AAY17087
 ID AAY17087 standard; protein; 639 AA.
 XX
 AC AAY17087;
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE An alkaline protease sequence from Bacillus species.
 XX
 KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent.
 XX
 OS Bacillus sp.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1. 639
 FT /note= "all residues indicated as Xaa are arbitrary amino
 FT acids"
 XX
 PN WO9918218-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX
 PR 07-OCT-1997; 97JP-00274570.
 XX
 PA (KAOS) KAO CORP.
 XX
 PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX
 DR WPI; 1999-287736/27.
 DR N-PSDB; AAX37277.
 XX
 PT Alkali protease from Bacillus used in washing powders.
 XX
 PS Claim 3; Page 47-50; 71pp; Japanese.
 XX
 CC The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease of the
 CC invention. (Updated on 20-MAR-2003 to correct DR field.)

```

XX SQ Sequence 639 AA;
Query Match 95.7%; Score 2150; DB 2; Length 639;
Best Local Similarity 96.1%; Pred. No. 4.4e-145;
Matches 417; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAGSSVGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 206 NDVARGIVKADVAGSSVGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 265
Qy 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
Db 266 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 325
Qy 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNPGNGGTTISAPGTAKNAI 180
Db 326 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNPGNGGTTISAPGTAKNAI 385
Qy 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKDPVMAPTFILSARSSLAPDSSF 240
Db 386 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKDPVMAPTFILSARSSLAPDSSF 445
Qy 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKALITAGAADIGLY 300
Db 446 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKALITAGAADIGLY 505
Qy 301 PNGNQGWGRVTLDKSLNVAAYVNESSLSSTQKATYSFTATAGKPLKISLVWSDAPASTTA 360
Db 506 PNGNQGWGRVTLDKSLNVAAYVNESSLSSTQKATYXFTATAGKPLKISLVWSDAPASTTA 565
Qy 361 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDWDGRNNVENVFNAPOSQGTYYTIEVQAYN 420
Db 566 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDWDGRNNVENVFNAPOSQGTYYTIEVQAYN 625
Qy 421 VPVGPQTFSLAIVN 434
Db 626 VPVGPQTFSLAIVN 639

RESULT 7
AAV17088
ID AAV17088 standard; protein; 640 AA.
XX AC AAV17088;
XX DT 20-MAR-2003 (revised)
XX DT 21-JUL-1999 (first entry)
XX DE An alkaline protease sequence from Bacillus species.
XX KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
XX KW washing composition; oxidising agent.
XX OS Bacillus sp.
XX FH Key Location/Qualifiers
XX FT Misc-difference 1..640
XX FT /note= "all residues indicated as Xaa are arbitrary amino
XX FT acids"
XX XX
XX PN WO9918218-A1.
XX XX
XX PD 15-APR-1999.
XX PF 07-OCT-1998; 98WO-JP004528.
XX PR 07-OCT-1997; 97JP-00274570.
XX PA (KAOS ) KAO CORP.
XX PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
XX PI Shikata S, Nomura M;

```

```

XX WPI: 1999-287736/27.
XX N-PSDB; AAX37278.
XX PT Alkali protease from Bacillus used in washing powders.
XX PS Claim 3; Page 50-53; 7ipp; Japanese.
XX CC The invention relates to alkaline proteases produced by strains of
XX CC Bacillus. The proteases ability to digest casein is not inhibited by
XX CC oleic acid and they have a high stability to oxidising agents. The
XX CC alkaline protease of the invention has the following properties: (a) it
XX CC is active over the pH range 4-13 and has at least 80% of its optimum
XX CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
XX CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
XX CC its ability to digest casein is not inhibited by oleic acid; (e) it has
XX CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
XX CC used as enzymes in washing compositions for use in automatic dishwashers
XX CC and for washing clothes. The stability to oxidising agents allows the
XX CC enzyme to be an effective component of washing compositions including
XX CC bleaches. The present sequence represents an alkaline protease of the
XX CC invention. (Updated on 20-MAR-2003 to correct DR field.)
XX SQ Sequence 640 AA;
Query Match 95.7%; Score 2150; DB 2; Length 640;
Best Local Similarity 96.1%; Pred. No. 4.5e-145;
Matches 417; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAGSSVGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 207 NDVARGIVKADVAGSSVGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266
Qy 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
Db 267 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 326
Qy 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNPGNGGTTISAPGTAKNAI 180
Db 327 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNPGNGGTTISAPGTAKNAI 386
Qy 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKDPVMAPTFILSARSSLAPDSSF 240
Db 387 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKDPVMAPTFILSARSSLAPDSSF 446
Qy 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKALITAGAADIGLY 300
Db 447 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKALITAGAADIGLY 506
Qy 301 PNGNQGWGRVTLDKSLNVAAYVNESSLSSTQKATYSFTATAGKPLKISLVWSDAPASTTA 360
Db 507 PNGNQGWGRVTLDKSLNVAAYVNESSLSSTQKATYXFTATAGKPLKISLVWSDAPASTTA 566
Qy 361 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDWDGRNNVENVFNAPOSQGTYYTIEVQAYN 420
Db 567 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDWDGRNNVENVFNAPOSQGTYYTIEVQAYN 626
Qy 421 VPVGPQTFSLAIVN 434
Db 627 VPVGPQTFSLAIVN 640

RESULT 8
AAM50085
ID AAM50085 standard; protein; 434 AA.
XX AC AAM50085;
XX AC AAM50085;
XX DT 12-AUG-2002 (first entry)
XX DE Bacillus sp alkaline protease protein A-1 fragment.
XX KW Alkaline protease; detergent; laundry; bleaching; dishwasher.

```

```

XX OS Bacillus sp.
XX PN EP1209233-A2.
XX PD 29-MAY-2002.
XX PF 22-NOV-2001; 2001EP-00127851.
XX PR 22-NOV-2000; 2000JP-00355166.
XX PR 12-APR-2001; 2001JP-00114048.
XX PA (KAOS ) KAO CORP.
XX PI Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
XX PI Okuda M, Saeki K;
XX DR WPI; 2002-437518/47.
XX PT New modified alkaline proteases useful in detergent compositions.
XX PS Claim 5; Page 18-19; 25pp; English.
XX CC This invention describes novel Bacillus sp. alkaline proteases useful in
XX CC detergent compositions, especially in laundry, bleaching or automatic
XX CC dishwasher detergents. The novel proteases have an increased detergency &
XX CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
XX CC sequence represents a fragment of the alkaline protease A-1 from Bacillus
XX CC sp NCIB12289 described in the method of the invention
XX SQ Sequence 434 AA;
Query Match 95.1%; Score 2138; DB 5; Length 434;
Best Local Similarity 93.3%; Pred. No. 1.9e-144;
Matches 405; Conservative 20; Mismatches 9; Indels 0; Gaps 0;
QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHAEFRGKITAIYALGRTN 60
DB 1 NDVARGIVKADVAQSSYGLYGQGVVAVADTGLDGRNDSMHAEFRGKITAIYALGRTN 60
QY 61 NANDTNGHGHVAGSVLGNSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NANDPNGHGHVAGSVLGNSTNKGMAPQANLVFQSVMDNSGLGGLPSNVSTLFSQAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 121 AGARIHTNSGAPVNGAYTTDSRNVDYVRKNDMAVLFQSVMDNSGLGGLPSNVSTLFSQAYS 180
QY 181 TVGATENLRPSFGSYADNINHVAQFSRGTQDGRIPKDVMAFGTIFLSARSSLAPDSGF 240
DB 181 TVGATENLRPSFGSYADNINHVAQFSRGTQDGRIPKDVMAFGTIFLSARSSLAPDSGF 240
QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
DB 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
QY 301 PNGNQGWGRVTLDKSLNVAVYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
DB 301 PNGNQGWGRVTLDKSLNVAVYVNESSLSSTLNQKATYSFTAQSGKPLKISLVWSDAPASTTA 360
QY 361 SVTLVNDLNLVITAPNGTQVGNDFTSFYNDNWDGRNNVENVFVINAPOSQGTYYIEVQAYN 420
DB 361 SVTLVNDLNLVITAPNGTKVGNDFTFAPYDNNWDGRNNVENVFVINAPOSQGTYYIEVQAYN 420
QY 421 VPVGQPTFSLAIVN 434
DB 421 VPQGPQAFSLAIVN 434
RESULT 9
AAMS0086
ID AAMS0086 standard; protein: 433 AA.
XX

```

```

AC AAM50086;
XX 12-AUG-2002 (first entry)
XX DE Bacillus sp alkaline protease protein A-2 fragment.
XX KW Alkaline protease; detergent; laundry; bleaching; dishwasher.
XX OS Bacillus sp.
XX PN EP1209233-A2.
XX PD 29-MAY-2002.
XX PF 22-NOV-2001; 2001EP-00127851.
XX PR 22-NOV-2000; 2000JP-00355166.
XX PR 12-APR-2001; 2001JP-00114048.
XX PA (KAOS ) KAO CORP.
XX PI Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
XX PI Okuda M, Saeki K;
XX DR WPI; 2002-437518/47.
XX PT New modified alkaline proteases useful in detergent compositions.
XX PS Claim 5; Page 20-21; 25pp; English.
XX CC This invention describes novel Bacillus sp. alkaline proteases useful in
XX CC detergent compositions, especially in laundry, bleaching or automatic
XX CC dishwasher detergents. The novel proteases have an increased detergency &
XX CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
XX CC sequence represents a fragment of the alkaline protease A-2 from Bacillus
XX CC sp NCIB12513 described in the method of the invention
XX SQ Sequence 433 AA;
Query Match 94.4%; Score 2120.5; DB 5; Length 433;
Best Local Similarity 93.3%; Pred. No. 3.4e-143;
Matches 405; Conservative 20; Mismatches 8; Indels 1; Gaps 1;
QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHAEFRGKITAIYALGRTN 60
DB 1 NDVARGIVKADVAQSNFGLYGQGIIVAVADTGLDGRNDSMHAEFRGKITAIYALGRTN 60
QY 61 NANDTNGHGHVAGSVLGNSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NANDPNGHGHVAGSVLGN-ATNKGMAPQANLVFQSIMDSGGGLGGLPANLQTLFSQAYS 119
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 120 AGARIHTNSGAPVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNINHVAQFSRGTQDGRIPKDVMAFGTIFLSARSSLAPDSGF 240
DB 180 TVGATENLRPSFGSYADNINHVAQFSRGTQDGRIPKDVMAFGTIFLSARSSLAPDSGF 239
QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
DB 240 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADVGLGF 299
QY 301 PNGNQGWGRVTLDKSLNVAVYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
DB 300 PNGNQGWGRVTLDKSLNVAVYVNETSPLSTOKATYSFTAQAGKPLKISLVSDAPGSTTA 359
QY 361 SVTLVNDLNLVITAPNGTQVGNDFTSFYNDNWDGRNNVENVFVINAPOSQGTYYIEVQAYN 420
DB 360 SLTLVNDLNLVITAPNGTKVGNDFTFAPYDNNWDGRNNVENVFVINAPOSQGTYYIEVQAYN 419
QY 421 VPVGQPTFSLAIVN 434

```

Db	420	VPVSPQTFSLAIVH 433	
Db	269	NANDPNGHGTAVGSLVGN-ATNKGMAPQANLVFQSIMDSGGGLGGLPANLQTLFSQAYS 327	
Db	121	AGARIHTNSWGAAYNGAYTTDSNRNVDYVRKNDMTILFAAGNEGPNNGTTISAPGTAKNAI 180	
Db	328	AGARIHTNSWGAAYNGAYTTDSNRNVDYVRKNDMTILFAAGNEGPNNGTTISAPGTAKNAI 387	
Db	181	TVGATENLRPSFGSYADNINHVAFQSSRGTPKGRIPKPDVMAFGTFTLSARSSILAPDSSF 240	
Db	388	TVGATENLRPSFGSYADNINHVAFQSSRGTPKGRIPKPDVMAFGTFTLSARSSILAPDSSF 447	
Db	241	WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKALIAAGADIGLGY 300	
Db	448	WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKALIAAGADIGLGY 507	
Db	301	PNGQGWGRVTLDKSLNVAVYNESSLSSTQKATYSFTATAGKPLKISLVNSDAPASTTA 360	
Db	508	PNGQGWGRVTLDKSLNVAVYNESSLSSTQKATYSFTATAGKPLKISLVNSDAPASTTA 567	
Db	361	SVTLVNDLNLVITAPNGTCVYVGNDFPTSPYNDNWDGRNNVFNAPQSGTYTIEVQAYN 420	
Db	568	SVTLVNDLNLVITAPNGTCVYVGNDFPTSPYNDNWDGRNNVFNAPQSGTYTIEVQAYN 627	
Db	421	VPVSPQTFSLAIVN 434	
Db	628	VPVSPQTFSLAIVH 641	
RESULT 11			
AAW50090	ID	AAW50090 standard; protein; 434 AA.	
XX	AC	AAW50090;	
XX	DT	12-AUG-2002 (first entry)	
XX	DE	Bacillus sp KSM-KP43 alkaline protease protein variant.	
XX	KW	Alkaline protease; detergent; laundry; bleaching; dishwasher; mutant;	
XX	KW	Mutain.	
XX	OS	Bacillus sp.	
XX	OS	Synthetic.	
XX	Key	Location/Qualifiers	
XX	FT	Misc-difference 46	/label= Y,w,a,d,e,t,v,l,i,h,s,k,q,m,c
XX	FT	/note= "as claimed in Claim 3"	
XX	FT	Misc-difference 54	/label= Y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c
XX	FT	/note= "as claimed in Claim 3"	
XX	FT	Misc-difference 57	/label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a
XX	FT	/note= "as claimed in Claim 3"	
XX	FT	Misc-difference 66	/label= e,d,s,q,a,t,l,m,c,v,g,i
XX	FT	/note= "as claimed in Claim 3"	
XX	FT	Misc-difference 84	/label= OTHER, R
XX	FT	/note= "OTHER= deleted residue. Specifically described in Claim 1"	
XX	FT	Misc-difference 101..106	/label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a
XX	FT	/note= "as claimed in Claim 3"	
XX	FT	Misc-difference 104	/label= OTHER, P
XX	FT	/note= "OTHER= deleted residue. Specifically described in Claim 1"	
XX	FT	Misc-difference 107	/label= k,r,a,s
XX	FT	/note= "as claimed in Claim 3"	
XX	FT	Misc-difference 119	/label= Y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c

Db	420	VPVSPQTFSLAIVH 433	
AAW89547	ID	AAW89547 standard; protein; 641 AA.	
XX	AC	AAW89547;	
XX	DT	12-APR-1999 (first entry)	
XX	DE	Bacillus JPI70 protease.	
XX	KW	Protease; detergent; surfactant; leather processing; debittering;	
XX	KW	flavour.	
XX	OS	Bacillus sp.	
XX	Key	Location/Qualifiers	
XX	FT	Peptide 1..33	/note= "signal peptide"
XX	FT	Region 34..208	/note= "prepro region"
XX	FT	Protein 209..641	/note= "mature protein"
XX	PN	WO9856927-A2.	
XX	PD	17-DEC-1998.	
XX	PF	09-JUN-1998; 98WO-US012005.	
XX	PR	12-JUN-1997; 97US-00873479.	
XX	PA	(NOVO) NOVO NORDISK BIOTECH INC.	
XX	PI	Sioma A, Christianson L;	
XX	DR	WPT; 1999-080908/07.	
XX	DR	N-PSDB; AAW82382.	
XX	FT	Novel protease from Bacillus subtilis LC20 - useful in laundry and	
XX	FT	dishwashing detergents and for leather processing.	
XX	PS	Claim 7; Page 53-54; 77pp; English.	
XX	CC	This is the amino acid sequence of a novel protease of Bacillus sp. JPI70	
XX	CC	(NCIB 12513), as deduced from the nucleotide sequence of an isolated gene	
XX	CC	(see AAW82382). The entire protein, including the signal peptide and	
XX	CC	prepro region, has 77% identity to alkaline protease Y (see AAW89548)	
XX	CC	from Bacillus. The invention provides vectors, recombinant host cells and	
XX	CC	methods for the recombinant production of the protease. The protease is	
XX	CC	used in laundry and dishwashing detergents, for institutional and	
XX	CC	industrial cleaning, and for leather processing, as well as for	
XX	CC	debittering and enhancing the degree of hydrolysis of protein	
XX	CC	hydrolyzates, for flavour development through hydrolysis of proteins,	
XX	CC	degradation of undesired peptides and in enzymatic synthesis of peptides.	
XX	CC	It has enhanced stability towards oxidation under alkaline conditions,	
XX	CC	e.g. towards bleaching agents of the peroxy type. The invention also	
XX	CC	provides mutant cells in which the protease activity is diminished. Such	
XX	CC	cells can be used for the production of heterologous recombinant proteins	
XX	SQ	Sequence 641 AA;	
Query Match		94.4%; Score 2120.5; DB 2; Length 641;	
Best Local Similarity		93.3%; Pred. No. 5.7e-143;	
Matches 405; Conservative		20; Mismatches 8; Indels 1; Gaps 1;	
Qy	1	NDVARGIVKADVAQSSVGLYCGQIVAVADTGTGNDSSMHEAFKITALYALGRTN 60	
Db	209	NDVARGIVKADVAQSSVGLYCGQIVAVADTGTGNDSSMHEAFKITALYALGRTN 268	
Qy	61	NANDTNGHGTAVGSLVGNSTNKGMAPQANLVFQSIMDSGGGLGGLPANLQTLFSQAYS 120	

FT Misc-difference 124 /note= "as claimed in Claim 3"

FT /label= a,k

FT /note= "as claimed in Claim 3"

FT Misc-difference 136 /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a

FT /note= "as claimed in Claim 3"

FT Misc-difference 138 /label= y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c

FT /note= "as claimed in Claim 3"

FT Misc-difference 148 /label= y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c

FT /note= "as claimed in Claim 3"

FT Misc-difference 193 /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a

FT /note= "as claimed in Claim 3"

FT Misc-difference 195 /label= y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c

FT /note= "as claimed in Claim 3"

FT Misc-difference 205 /label= y,w,a,d,e,t,v,l,i,h,s,k,q,m,c

FT /note= "as claimed in Claim 3"

FT Misc-difference 247 /label= w,f,a,r,e,t,v,l,i,h,s,e,m,c

FT /note= "as claimed in Claim 3"

FT Misc-difference 256 /label= OTHER, a,s,e,v,l,r,e,d

FT /note= "OTHER= deleted residue. Specifically described in Claim 1"

FT Misc-difference 257 /label= v,i

FT /note= "as claimed in Claim 3"

FT Misc-difference 264 /label= e,d,s,g,a,t,l,m,c,v,g,i

FT /note= "as claimed in Claim 3"

FT Misc-difference 342 /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a

FT /note= "as claimed in Claim 3"

FT Misc-difference 369 /label= OTHER, d

FT /note= "OTHER= deleted residue. Specifically described in Claim 1"

FT

PN EP1209233-A2.

XX

XX 29-MAY-2002.

XX

XX 22-NOV-2001; 2001EP-00127851.

XX

XX 22-NOV-2000; 2000JP-00355166.

XX 12-APR-2001; 2001JP-00114048.

XX

XX (KAOS) KAO CORP.

XX Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;

PI Okuda M, Saeki K;

XX WPI; 2002-437518/47.

XX New modified alkaline proteases useful in detergent compositions.

XX Claim 1; Page; 25pp; English.

XX

XX This invention describes novel Bacillus sp. alkaline proteases useful in detergent compositions, especially in laundry, bleaching or automatic dishwasher detergents. The novel proteases have an increased detergency (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This sequence represents a variant of the alkaline protease KP43 from Bacillus sp strain KSM-KP43 created from the wild-type protease represented in

CC AAM50080 NOTE: This sequence is not represented in the specification but has been constructed from the sequence represented in record AAM50080

XX

XX Sequence 434 AA;

Query Match 92.2%; Score 2071; DB 5; Length 434;
Best Local Similarity 94.0%; Pred. No. 1.2e-139;
Matches 408; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFPRGKITALYALGRIN 60
DB 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAXRGKITALXALXRN 60
QY 61 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIIMDSGGGLGGLPSNLQTLFSQAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILPAAGNEGPNNGGTISAPGTAKNAI 180
DB 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILPAAGNEGPNNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAQSFSSRGPTKDGRIKPDVMAFGPTFIISARSLAPDSF 240
DB 181 TVGATENLRPSFGSYADNINHVAQSFSSRGPTKDGRIKPDVMAFGPTFIISARSLAPDSF 240
QY 241 WANHDSKIYMGTSMTATPIVAGNVAQLREHFVKRGITPKSLLKALIAAGADIGLCY 300
DB 241 WANHDSKIYMGTSMTATPIVAGNVAQLREHFVKRGITPKSLLKALIAAGADIGLCY 300
QY 301 PNGNQGWGRVTLDKSLNVAIVNNESSLSSTSQATYSFTATAGKPLKISLVMSDAPASTTA 360
DB 301 PNGNQGWGRVTLDKSLNVAIVNNESSLSSTSQATYSFTATAGKPLKISLVMSDAPASTTA 360
QY 361 SVTLVNDLNLVITAPNGTQVGNDFTPYNDNWDGNNVENYFINAPQSGTYTIEQAYN 420
DB 361 SVTLVNDLNLVITAPNGTQVGNDFTPYNDNWDGNNVENYFINAPQSGTYTIEQAYN 420
QY 421 VPVGPQTFSLAIYN 434
DB 421 VPVGPQTFSLAIYN 434

RESULT 12
AAM50084
ID AAM50084 standard; protein; 433 AA.
XX
AC AAM50084;
XX
XX 12-AUG-2002 (first entry)
XX
DE Bacillus sp SD-521 (FERM BP-11162) alkaline protease protein fragment.
XX
XX Alkaline protease; detergent; laundry; bleaching; dishwasher.
XX
XX Bacillus sp.
XX
XX EP1209233-A2.
XX
XX 29-MAY-2002.
XX
XX 22-NOV-2001; 2001EP-00127851.
XX
XX 22-NOV-2000; 2000JP-00355166.
XX 12-APR-2001; 2001JP-00114048.
XX
XX (KAOS) KAO CORP.
XX
XX Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
PI Okuda M, Saeki K;
XX
XX WPI; 2002-437518/47.
XX
XX New modified alkaline proteases useful in detergent compositions.
XX
XX Claim 5; Page 16-18; 25pp; English.
XX
XX This invention describes novel Bacillus sp. alkaline proteases useful in

XX 12-JUN-1997; 97US-00873479.
 PR (NOVO) NOVO NORDISK BIOTECH INC.
 XX
 PA Sloma A, Christianson L;
 XX
 PI WPI; 1999-080908/07.
 XX
 DR Novel protease from *Bacillus subtilis* LC20 - useful in laundry and
 PT dishwashing detergents and for leather processing.
 PT
 XX Claim 3; Page 55-56; 77pp; English.
 PS
 XX This is the amino acid sequence of a *Bacillus* sp. alkaline protease Y
 CC that is said to have good alkali and surfactant resistance and improved
 CC detergency. It shows 77% identity to a newly isolated protease (see
 CC AA089547) of *Bacillus* sp. JP170 (NCIB 12513). The invention provides
 CC vectors, recombinant host cells and methods for the recombinant
 CC production of such proteases. The protease are used in laundry and
 CC dishwashing detergents, for institutional and industrial cleaning, and
 CC for leather processing, as well as for debittering and enhancing the
 CC degree of hydrolysis of protein hydrolysates, for flavour development
 CC through hydrolysis of proteins, degradation of undesired peptides and in
 CC enzymatic synthesis of peptides. They have enhanced stability towards
 CC oxidation under alkaline conditions, e.g. towards bleaching agents of the
 CC peroxy type. The invention also provides mutant cells in which the
 CC protease activity is diminished. Such cells can be used for the
 CC production of heterologous recombinant proteins
 XX
 XX Sequence 636 AA;
 SQ
 Query Match 88.3%; Score 1984.5; DB 2; Length 636;
 Best Local Similarity 87.1%; Pred. No. 2.9e-133;
 Matches 378; Conservative 31; Mismatches 24; Indels 1; Gaps 1;
 QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
 DB 204 NDVARGIVKADVAQNYGLYGQQLVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 263
 QY 61 NANTDNGHGHVAGSVLNGSTKNGAPQANLVFQSIIMDSGGGLGGLPSNLTLFSQAYS 120
 DB 264 NASDPNGHGHVAGSVLGN-ALNKGAPQANLVFQSIIMDSGGGLGGLPSNLTLFSQAWN 322
 QY 121 AGARIHTNSGAAVNGAYTTDSNRVDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
 DB 323 AGARIHTNSGAPVNGAYTANSRQVDEYVRNNDMTVLFAAGNEGPNNGTISAPGTAKNAI 382
 QY 181 TVGATENLRPSFGSIADNPNHIAQFSSRGATRDGRIPKDVTAFTGTFILSARSLAPDSSF 240
 DB 383 TVGATENLRPSFGSIADNPNHIAQFSSRGATRDGRIPKDVTAFTGTFILSARSLAPDSSF 442
 QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALIAGAADIGLY 300
 DB 443 WANYSKYAYMGGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALIAGAADIGLY 502
 QY 301 PNGNQGWGRVTLDKSLNVAVYVNSSLSSTQKATYSTATAGPKLKI SLVWSDAPASTTA 360
 DB 503 PSQDQGWGRVTLDKSLNVAVYVNEATALATGQKATYSFQAQKPLKISLVWTDAPGSTTA 562
 QY 361 SVTLVNDLNLVITAPNGTQVYVNDFTSPYNDNDGNNVNFVFNAPQSGTYYTIEVOAYN 420
 DB 563 SYTLVNDLNLVITAPNGQKYVGNDFSAFYDNNWDGNNVNFVFNAPQSGTYYTIEVOAYN 622
 QY 421 VPVGPQTFFSLAIVN 434
 DB 623 VPSPGFQFSLAIVH 636
 RESULT 15
 AAM50083
 ID AAM50083 standard; protein; 433 AA.
 XX

AC AAM50083;
 XX 12-AUG-2002 (first entry)
 DT
 XX *Bacillus* sp Y-(FERM BP-1029) alkaline protease protein fragment.
 DE
 XX Alkaline protease; detergent; laundry; bleaching; dishwasher.
 KW
 XX *Bacillus* sp.
 OS
 XX EPI2092333-A2.
 PN
 XX 29-MAY-2002.
 PD
 XX 22-NOV-2001; 2001EP-00127851.
 PF
 XX 22-NOV-2000; 2000JP-00355166.
 PR
 XX 12-APR-2001; 2001JP-00114048.
 XX (KAOS) KAO CORP.
 PA Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
 PI Okuda M, Saeki K;
 FI
 XX WPI; 2002-437518/47.
 DR
 XX New modified alkaline proteases useful in detergent compositions.
 PT
 XX Claim 5; Page 15-16; 25pp; English.
 PS
 XX This invention describes novel *Bacillus* sp. alkaline proteases useful in
 CC detergent compositions, especially in laundry, bleaching or automatic
 CC dishwasher detergents. The novel proteases have an increased detergency &
 CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
 CC sequence represents a fragment of the alkaline protease Ya from *Bacillus*
 CC sp strain Y-(FERM BP-1029) described in the method of the invention
 XX
 XX Sequence 433 AA;
 SQ
 Query Match 88.2%; Score 1982.5; DB 5; Length 433;
 Best Local Similarity 87.1%; Pred. No. 2.5e-133;
 Matches 378; Conservative 30; Mismatches 25; Indels 1; Gaps 1;
 QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
 DB 1 NDVARGIVKADVAQNYGLYGQQLVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
 QY 61 NANTDNGHGHVAGSVLNGSTKNGAPQANLVFQSIIMDSGGGLGGLPSNLTLFSQAYS 120
 DB 61 NASDPNGHGHVAGSVLGN-ALNKGAPQANLVFQSIIMDSGGGLGGLPSNLTLFSQAWN 119
 QY 121 AGARIHTNSGAAVNGAYTTDSNRVDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
 DB 120 AGARIHTNSGAPVNGAYTANSRQVDEYVRNNDMTVLFAAGNEGPNNGTISAPGTAKNAI 179
 QY 181 TVGATENLRPSFGSIADNPNHIAQFSSRGATRDGRIPKDVTAFTGTFILSARSLAPDSSF 240
 DB 180 TVGATENLRPSFGSIADNPNHIAQFSSRGATRDGRIPKDVTAFTGTFILSARSLAPDSSF 239
 QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALIAGAADIGLY 300
 DB 240 WANYSKYAYMGGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALIAGAADIGLY 299
 QY 301 PNGNQGWGRVTLDKSLNVAVYVNSSLSSTQKATYSTATAGPKLKI SLVWSDAPASTTA 360
 DB 300 PNGDQGWGRVTLNKSLNVAVYVNEATALATGQKATYSFQAQKPLKISLVWTDAPGSTTA 359
 QY 361 SVTLVNDLNLVITAPNGTQVYVNDFTSPYNDNDGNNVNFVFNAPQSGTYYTIEVOAYN 420
 DB 360 SYTLVNDLNLVITAPNGQKYVGNDFSAFYDNNWDGNNVNFVFNAPQSGTYYTIEVOAYN 419
 QY 421 VPVGPQTFFSLAIVN 434
 XX

Db 420 VPSGPQRFSLAIVH 433

Search completed: March 10, 2004, 14:53:09
Job time : 60 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:52:04 ; Search time 23 Seconds
(without alignments)
974.160 Million cell updates/sec

Title: US-09-985-689a-1-COPY
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAVNPVGPOTFSLAIWN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/PTUS COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfile1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2242	99.8	640	4	US-09-509-814A-6
2	2237	99.8	640	4	US-09-509-814A-8
3	2178	96.9	639	4	US-09-509-814A-4
4	2150	95.7	639	4	US-09-509-814A-1
5	2150	95.7	640	4	US-09-509-814A-2
6	2120.5	94.4	641	2	US-08-873-479-42
7	1981.5	88.2	433	4	US-09-104-623A-4
8	1981.5	88.2	433	4	US-09-019-532-4
9	1981.5	88.2	433	4	US-08-338-746-4
10	1981.5	88.2	435	2	US-08-873-479-43
11	1581.5	70.4	345	4	US-08-512-251A-10
12	1581.5	70.4	345	4	US-09-515-150A-10
13	1581.5	70.4	345	4	US-09-196-281-13
14	447.5	19.9	659	3	US-08-894-818B-1
15	447.5	19.9	659	4	US-09-445-472-12
16	409	18.2	412	4	US-08-445-472-1
17	409	18.2	522	3	US-08-894-818B-3
18	409	18.2	522	4	US-08-445-472-4
19	409	18.2	654	3	US-08-894-818B-35
20	409	18.2	654	4	US-09-445-472-16
21	396	17.6	659	3	US-08-894-818B-5
22	348	15.5	520	3	US-09-000-016-7
23	348	15.5	520	4	US-09-514-340-7
24	348	15.5	734	3	US-09-000-016-4
25	348	15.5	734	4	US-09-514-340-4
26	348	15.5	823	3	US-09-000-016-2
27	348	15.5	823	4	US-09-514-340-2

28	305.5	13.6	903	1	US-08-750-532-1	Sequence 1, Appli
29	305.5	13.6	1398	1	US-08-750-532-9	Sequence 9, Appli
30	305.5	13.6	1398	3	US-08-894-818B-8	Sequence 8, Appli
31	305.5	13.6	1398	4	US-09-445-472-6	Sequence 6, Appli
32	295.5	13.2	237	1	US-08-750-532-18	Sequence 18, Appli
33	283.5	12.6	418	2	US-08-873-479-44	Sequence 44, Appli
34	276	12.3	418	4	US-09-966-921A-2	Sequence 2, Appli
35	256.5	11.4	397	4	US-09-328-352-7533	Sequence 7533, Ap
36	254.5	11.3	275	1	US-08-431-387-1	Sequence 1, Appli
37	251	11.2	280	1	US-08-434-355-8	Sequence 8, Appli
38	251	11.2	280	1	US-08-459-967-8	Sequence 8, Appli
39	251	11.2	280	1	US-08-460-327-8	Sequence 8, Appli
40	251	11.2	280	1	US-08-459-871-8	Sequence 8, Appli
41	251	11.2	280	3	US-09-024-532-2	Sequence 2, Appli
42	251	11.2	280	4	US-09-104-623A-2	Sequence 2, Appli
43	251	11.2	280	4	US-09-019-532-2	Sequence 2, Appli
44	251	11.2	280	4	US-09-417-359A-2	Sequence 2, Appli
45	251	11.2	280	4	US-09-705-185-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-509-814A-6
; Sequence 6, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUMI
; APPLICANT: NOMURA, MASARUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-OPCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 6
; LENGTH: 640
; TYPE: PRT
; ORGANISM: Bacillus sp.
; US-09-509-814A-6

Query Match	99.8%	Score	2242	DB 4	Length	640	
Best Local Similarity	99.8%	Pred. No.	7.4e-172				
Matches	433	Conservative	1	Mismatches	0	Gaps	0
QY	1	NDVARGIVKADVAQSSYGLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRN	60				
Db	207	NDVARGIVKADVAQSSYGLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRN	266				
QY	61	NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIIMDSGGGLGGLPSNLQTLFSQAYS	120				
Db	267	NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIIMDSGGGLGGLPSNLQTLFSQAYS	326				
QY	121	AGARIHTNSGVAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI	180				
Db	327	AGARIHTNSGVAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI	386				
QY	181	TVGATENLRPSFGSYADNINHVAQFSRRGPTKDGRIKPDVMAFGTIFLSARSSLA	240				
Db	387	TVGATENLRPSFGSYADNINHVAQFSRRGPTKDGRIKPDVMAFGTIFLSARSSLA	446				
QY	241	WANHDSKIYAMGGTSMATPIVAGNVAQLREHFVKRGITPKPSLLKALIAAGAA	300				

Db 447 WANHDSKYAYMGSTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAGAADIIGLY 506
Qy 301 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTSQKATYSFTATAGKPKLSLWSDAPASTTA 360
Db 507 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTSQKATYSFTATAGKPKLSLWSDAPASTTA 566
Qy 361 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDNWDGRNNVNFVFINAPQSGTYTIEVQAYN 420
Db 567 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDNWDGRNNVNFVFINAPQSGTYTIEVQAYN 626
Qy 421 VPVGPQTFSLAIVN 434
Db 627 VPVGPQTFSLAIVN 640

RESULT 2

US-09-509-814A-8
; Sequence 8, Application US/09509814A

; Patent No. 6376227

; GENERAL INFORMATION:

; APPLICANT: TAKAIWA, MIKIO

; APPLICANT: OKUDA, MITSUYOSHI

; APPLICANT: SAEKI, KATSUHIISA

; APPLICANT: KUBOTA, HIROMI

; APPLICANT: HITOMI, JUN

; APPLICANT: KAGEYAMA, YASUSHI

; APPLICANT: SHIKATA, SHITSUM

; APPLICANT: NOMURA, MASAFUMI

; TITLE OF INVENTION: ALKALINE PROTEASE

; FILE REFERENCE: 0327-0832-OPCT

; CURRENT APPLICATION NUMBER: US/09/509,814A

; CURRENT FILING DATE: 2000-04-06

; PRIOR APPLICATION NUMBER: PCT/JP98/04528

; PRIOR FILING DATE: 1998-10-07

; PRIOR APPLICATION NUMBER: JP 9-274570

; PRIOR FILING DATE: 1997-06-08

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 8

; LENGTH: 640

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-509-814A-8

Query Match 99.6%; Score 2237; DB 4; Length 640;

Best Local Similarity 99.5%; Pred. No. 1.9e-171;

Matches 432; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 207 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266
Qy 61 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLOTLSOAYS 120
Db 267 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLOTLSOAYS 326
Qy 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 327 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
Qy 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTILSARSLAPDSF 240
Db 387 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTILSARSLAPDSF 446
Qy 241 WANHDSKYAYMGSTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAGAADIIGLY 300
Db 447 WANHDSKYAYMGSTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAGAADIIGLY 506
Qy 301 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTSQKATYSFTATAGKPKLSLWSDAPASTTA 360
Db 507 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTSQKATYSFTATAGKPKLSLWSDAPASTTA 566

Qy 361 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDNWDGRNNVNFVFINAPQSGTYTIEVQAYN 420
Db 567 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDNWDGRNNVNFVFINAPQSGTYTIEVQAYN 626
Qy 421 VPVGPQTFSLAIVN 434
Db 627 VPVGPQTFSLAIVN 640

RESULT 3

US-09-509-814A-4

; Sequence 4, Application US/09509814A

; Patent No. 6376227

; GENERAL INFORMATION:

; APPLICANT: TAKAIWA, MIKIO

; APPLICANT: OKUDA, MITSUYOSHI

; APPLICANT: SAEKI, KATSUHIISA

; APPLICANT: KUBOTA, HIROMI

; APPLICANT: HITOMI, JUN

; APPLICANT: KAGEYAMA, YASUSHI

; APPLICANT: SHIKATA, SHITSUM

; APPLICANT: NOMURA, MASAFUMI

; TITLE OF INVENTION: ALKALINE PROTEASE

; FILE REFERENCE: 0327-0832-OPCT

; CURRENT APPLICATION NUMBER: US/09/509,814A

; CURRENT FILING DATE: 2000-04-06

; PRIOR APPLICATION NUMBER: PCT/JP98/04528

; PRIOR FILING DATE: 1998-10-07

; PRIOR APPLICATION NUMBER: JP 9-274570

; PRIOR FILING DATE: 1997-06-08

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 4

; LENGTH: 639

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-509-814A-4

Query Match 96.9%; Score 2178; DB 4; Length 639;

Best Local Similarity 96.1%; Pred. No. 1e-166;

Matches 417; Conservative 14; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 206 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 265
Qy 61 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLOTLSOAYS 120
Db 266 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLOTLSOAYS 325
Qy 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 326 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 385
Qy 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTILSARSLAPDSF 240
Db 386 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTILSARSLAPDSF 445
Qy 241 WANHDSKYAYMGSTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAGAADIIGLY 300
Db 446 WANHDSKYAYMGSTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAGAADIIGLY 505
Qy 301 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTSQKATYSFTATAGKPKLSLWSDAPASTTA 360
Db 506 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTSQKATYSFTATAGKPKLSLWSDAPASTTA 565
Qy 361 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDNWDGRNNVNFVFINAPQSGTYTIEVQAYN 420
Db 566 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDNWDGRNNVNFVFINAPQSGTYTIEVQAYN 625
Qy 421 VPVGPQTFSLAIVN 434
Db 626 VPVGPQTFSLAIVN 639